Synthesis and studies of oligonucleotides containing mutagenic 2'-deoxyguanosine derivatives.

James B. Thomson

A thesis submitted for the degree

of Doctor of Philosophy.

University of Edinburgh November 1992

This thesis is submitted in part fulfilment of the requirements of the degree of Doctor of Philosophy at Edinburgh University. Unless otherwise stated, the work described is original and has not been previously submitted, in whole or in part, for any degree at this or any other university.

University of Edinburgh

November 1992

"Facts are cheels that winna ding,

An' downa be disputed"

(Robert Burns, A Dream)

Acknowledgements.

I would like to thank Dr. T. Brown for the provision of research facilities and to gratefully acknowledge his enthusiasm and advice throughout the course of this work. My thanks also go to Dr. W. P. Watson of Shell Research who, in conjunction with the Science and Engineering Research Council have provided the funding for my research. I am very grateful for the facilities of Oswel, where all the oligonucleotides were synthesised, and to the support of the departmental technical services, in particular those responsible for the running of the NMR and mass spectrum analysis. Thanks also go to my friends and colleagues with whom I have enjoyed working with, in particular, Dr. G. Leonard for solving the crystal structure, Calumn MacKellar, Katherine McAuley and Catriona Tedford for their help and to James Macdonald for his optimism and wit.

Finally, I am indebted to my parents, Helen and John, for their continued encouragement and financial support over the years which has enabled me to pursue my studies and to my wife, Ingrid, for her thoughtfulness and patience. It is to her that this thesis is dedicated.

Abstract.

The self-complementary dodecanucleotide d[CGC$\underline{O^6MeG}$AATT$\underline{T}$GCG)$_2$ which contains two O^6-methyl-2'-deoxyguanosine-thymine base pairs, has been analysed by X-ray diffraction methods and the structure refined to a residual error of R=0.185 at 2.0*A* resolution. The O^6MedG.T mispair closely resembles a Watson / Crick base pair and there are very few structural differences between the O^6MedG.T duplex and the native analogue. The similarity between the O^6MedG.T base pair and a normal dG.dC base pair explains the failure of mismatch repair enzymes to recognise and remove this mutagenic lesion. A series of ultraviolet melting studies over a wide pH range on a related dodecamer indicate that the O^6MedG.dC mispair can exist in two conformations, one of which is a wobble pair, and the other a protonated Watson / Crick pair. The former, which predominates at physiological pH, will be removed by normal proofreading and repair enzymes, whereas the latter is likely to escape detection. Hence, the occasional occurence of the protonated O^6MedG.dC base pair may explain why the presence of O^6MedG in genomic DNA does not always give rise to a mutation.

Synthesis of C(8) substituted 2'-deoxyguanosine derivatives was also undertaken in an effort to yield 8-oxo-7-hydro-2'-deoxyguanosine, a product of oxidative damage to DNA. Displacement of bromine, from 8-bromo-2'-deoxyguanosine, was achieved with methoxide, benzyloxide and p-nitrothiophenol. Conversion of 8-methoxy-2'-deoxyguanosine to 8-oxo-7-hydro-2'-deoxyguanosine was achieved by use of thiophenol, however conversion was not possible using oligonucleotides containing the 8MeOdG nucleoside. Oligonucleotides containing 8-bromo-2'-deoxyguanosine were synthesised which demonstrated similar pairing properties to oligonucleotides containing 2'-deoxyguanosine. As a result 8BrdG was utilised as a heavy atom marker in the crystal structure of the DNA drug complex of d[T$\underline{8BrG}$TA$\underline{C}$A] / nogalomycin, thus aiding the refinement of the native DNA / drug sequence.

Abbreviations.

8BnOdG	8-benzyloxy-2'-deoxyguanosine.
8BrdA	8-bromo-2'-deoxyadenosine.
8BrdG	8-bromo-2'-deoxyguanosine.
8MeOdG	8-methoxy-2'-deoxyguanosine.
8NPTdG	8-(p-nitrophenylthio)-2'-deoxyguanosine.
8OxodG	8-oxo-7-hydro-2'-deoxyguanosine.
Ac	acetyl.
CPG	controlled pore glass.
dA	2'-deoxyadenosine.
DBU	diazabicyclo-[5. 4. 0.]-undec-7-ene.
dC	2'-deoxycytidine.
DCM	dichloromethane.
dec.	decomposes.
dG	2'-deoxyguanosine.
DMAP	4-dimethylaminopyridine.
DMF	dimethylformamide.
DMSO	dimethylsulphoxide.
DMTr	4,4'-dimethoxytrityl.
DMTrCl	4,4'-dimethoxytritylchloride.
dR	2'-deoxyribose.
EDTA	ethylene diamine tetraacetic acid.
FAB-MS	Fast Atom Bombardment Mass Spectrometry.
HPLC	High Performance Liquid Chromatography.
Ib	isobutyryl.

ip	isopropyl
mCPA	2-cyanoethyl-N-N'-diisopropylphosphor-amidochloridite.
n.m.r.	Nuclear Magnetic Resonance.

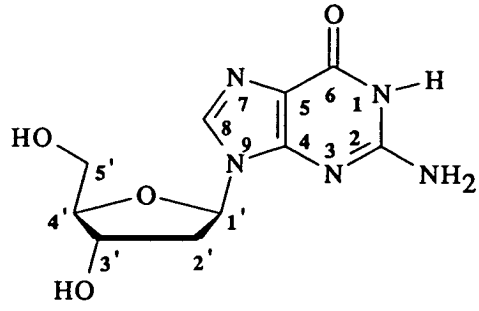
n.m.r. terminology.

br.	broad.
d.	doublet.
dd.	doublet of doublets.
exch.	exchangeable.
m.	multiplet.
qd.	quartet of doublets.
quin.	quintet.
sept.	septet.
sext.	sextet.
t.	triplet.

NPT	p-nitrophenylthio.
NTP	p-nitrothiophenol.
O^6MedG	O^6methyl-2'-deoxyguanosine.
PA	phenylacetyl.
psi.	pounds per square inch.
R_f.	Relative front.
Sh.	shoulder.
T	thymidine.
TEA	triethylamine.
T_m	Melting temperature.
$O^6MedG^{N^2Ib}$	O^6-methyl-N^2-isobutyryl-2'-deoxyguanosine.

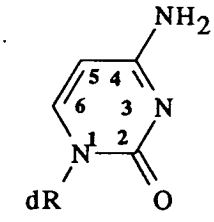
Ring numbering.

Purines.

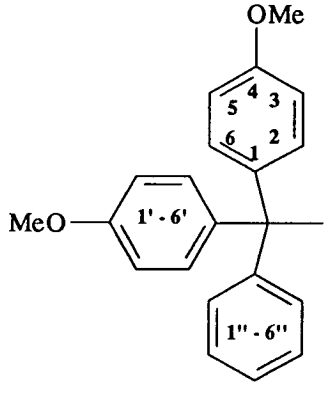


2'-deoxyguanosine.

Pyrimidines.



2'-deoxycytidine.



4,4'-dimethoxytrityl.

d[CGCGAATTCGCG] Oligonucleotides.
1 12.

Single stranded DNA is always written and numbered from the 5' to the 3' direction. For example, in the above sequence dC, the 5' nucleotide, is numbered one and dG, the 3' nucleotide, numbered twelve.

Contents.

INTRODUCTION.

1.1. Introduction to DNA.

The extreme diversity of plant and animal life existing on planet earth is a result of evolutionary processes stretching back over hundreds of millions of years. The conditions that were present in the prebiotic world are not known and many ideas have been put forward. One of the earliest attempts to mimic prebiotic conditions was the Miller experiment, (Miller 1953), which demonstrated the synthesis of amino acids from simple molecules, and more recently the work of Joyce (Joyce 1989) who attempted to understand the very complex series of reactions that would be necessary for the synthesis of ribonucleotides from simple molecules. It is now generally accepted that the process of evolution began with molecules that had the ability to act as templates for their own synthesis. Such molecules must also have possesed the ability to mutate and therefore adapt to their surroundings. Molecules with this ability are not easy to visualise, but some elegant work (Hong 1992) is beginning to provide an understanding of the potential of a prebiotic environment. The types of compounds present, are of course unknown, but it is clear today that ribo- and deoxyribo- nucleic acids (RNA and DNA) are the only classes of compounds which have continued to prosper from such elementary beginnings. They are present in every single cell of all known organisms. The discovery of catalytic RNA (Cech 1990 and Altman 1990) has added weight to the argument that RNA may have been the original replicating catalyst which was eventually superceded by deoxyribonucleic acids (to store the information) and proteins (to catalyse chemical reactions).

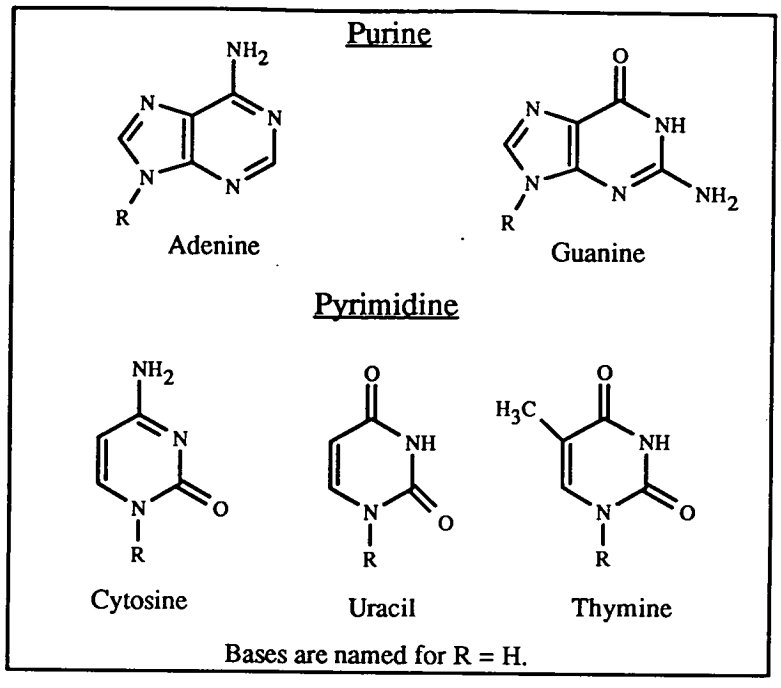
It had been known ever since Mendel (1822-1884) crossed round and wrinkled peas (Knowles, F.G.W. 1945), that there was some hidden force which resulted in

parental characteristics being inherited by the next generation. It was Avery, in 1944, who established the role of DNA in this process. Establishing exactly how this information is securely stored and passed on to subsequent generations is not a trivial problem, even when one considers a simple bacterium such as *E. Coli* which is unicellular and will, in ideal conditions, reproduce through cell division approximately every thirty minutes. This not only requires very rapid replication of the genome, such that both daughter cells have a copy, but also requires easy access to the DNA molecule which, by design, has a very compact and enclosed structure in order to safely protect the genetic material from any damaging agents and to package the genetic information into a small volume within the cell.

As far back as 1869, when nucleic acids were first isolated by Miescher, it was known that they were major constituents of cells and nuclei, hence the name "nucleic" (from the nucleus). Subsequent characterisation determined the presence of five different heterocyclic bases, two types of sugar rings and a phosphate group. Together these three subunits form the monomer nucleotide units (Figure 1.1a) which, when condensed together, (Figure 1.1b), create the polymeric structures of DNA and RNA. One of the biggest stumbling blocks in establishing the structure of DNA was that the correct tautomeric forms of the four heterocyclic bases A, G, C and T were not known. However the fact that in every DNA sequence analysed the A:T and G:C ratios were always around 1:1 [Chargaff's law (A+G)=(T+C)] suggested a very special relationship between the purine and pyrimidine nucleotides (Chargaff 1950). With aid of the fibre diffraction data of Franklin and Wilkins, Crick and Watson finally analysed all the available data and correctly determined the structure of B-DNA, publishing their findings in *Nature* (Watson and Crick 1953). By recognising the helical symmetry of the molecule they were able to establish that the bases, in their correct tautomeric forms, hydrogen bonded in a pseudo symmetric manner,

Figure 1.1a.

Base = B =



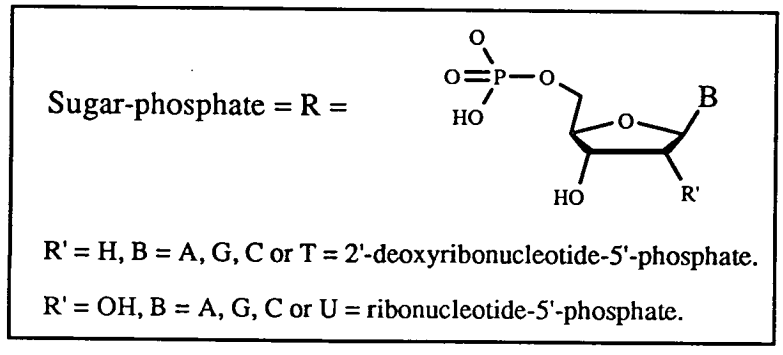
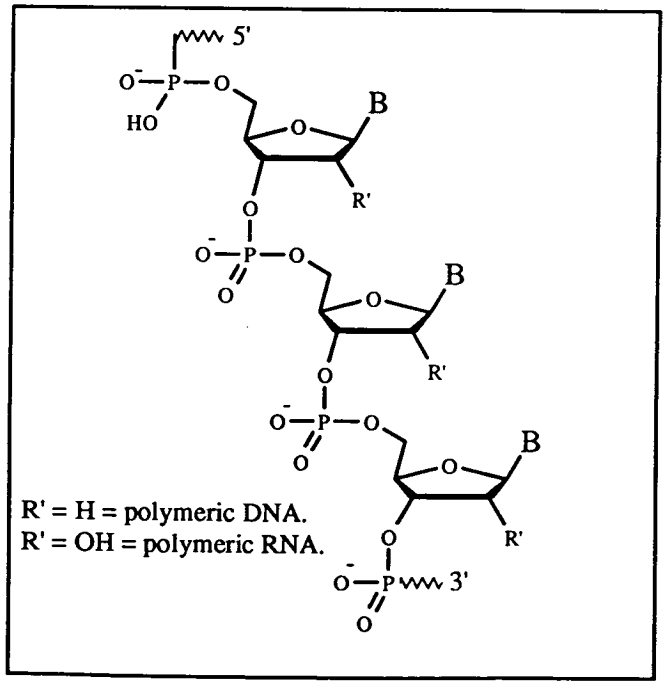
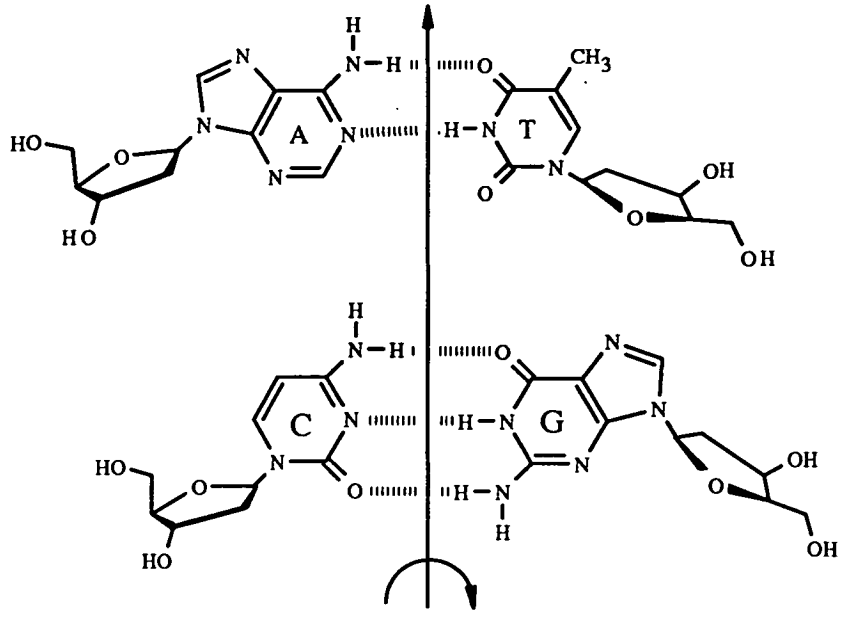


Figure 1.1b.



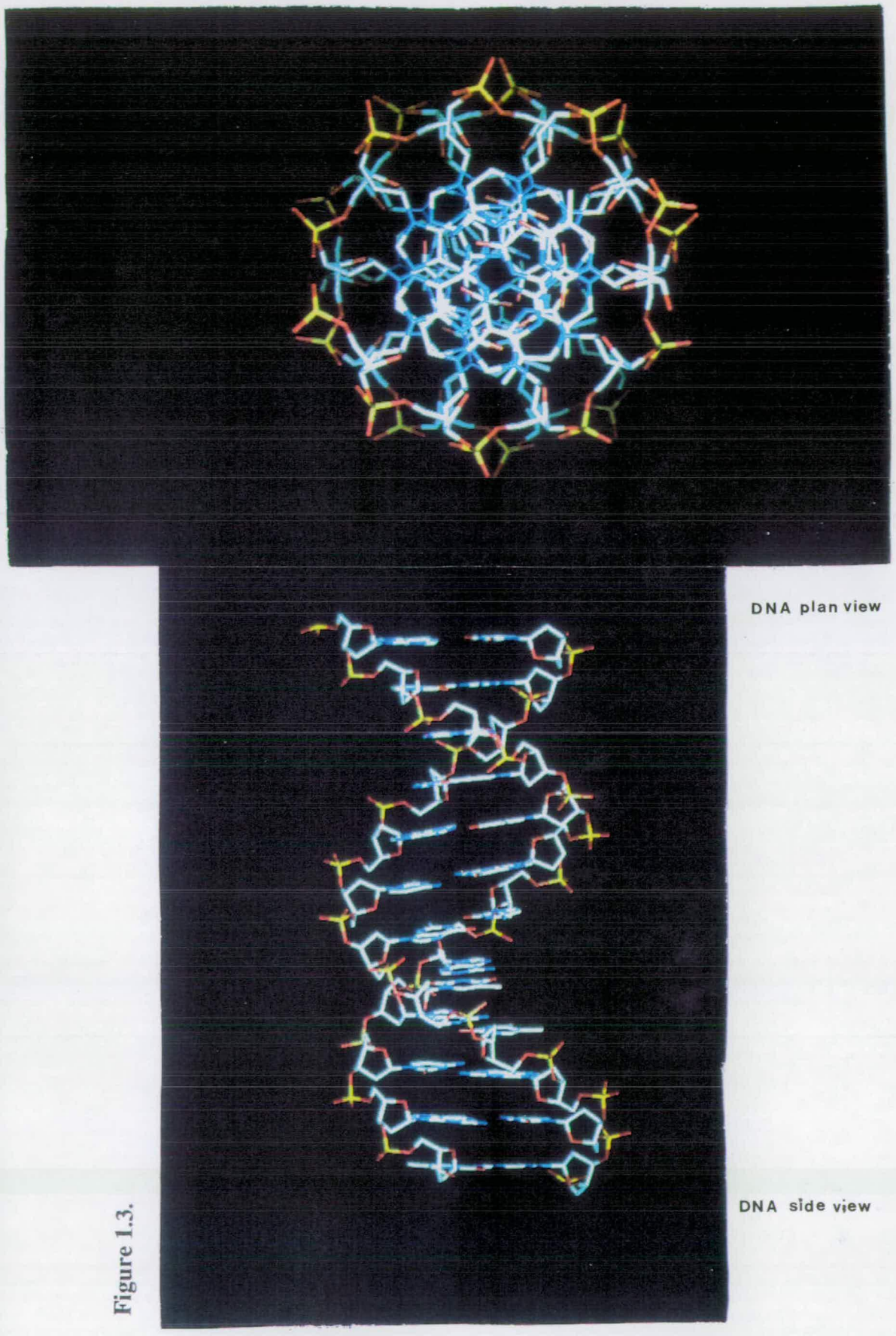
(Figure 1.2), thus explaining the rationale behind Chargaffs law.

Figure 1.2.



= Pseudodyad relating glycosidic bonds.

The structure of DNA consists of two strands which wind round each other in an antiparallel manner to yield a double helix with complementary hydrogen bonding holding the two strands together. The elegant symmetry and compactness of B-DNA can be seen from the various views, (Figure 1.3), which illustrate how the base pairs stack at the centre of the helix with the two phosphodiester-sugar backbones winding around the outside. This gives rise to two continuous grooves in the duplex which, owing to their different dimensions, are termed the major and minor grooves. Hydrogen bonding complementarity between the four bases was the crucial discovery which led to the structural determination which established that only A.T, T.A, G.C and C.G base pairs were possible. This not only explained the reasoning behind Chargaffs' law but also suggested a mechanism for replication. By storing all the information on only one strand the



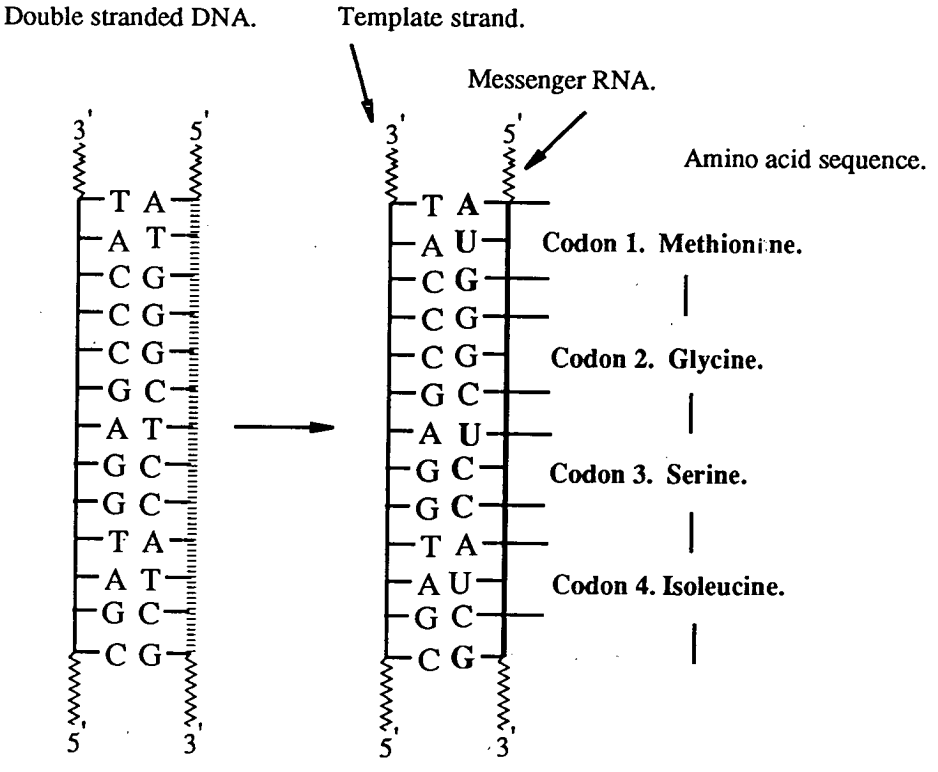
DNA plan view

DNA side view

Figure 1.3.

helix can unwind allowing each strand to act as a template for the synthesis of its complementary partner thus giving rise to two daughter copies of the original duplex. In 1968, Nirenberg, Ochoa and Khorana determined that DNA directed protein synthesis via a triplet code, with each block of three bases, called a codon, representing one amino acid. Thus the DNA sequence of a gene determines the amino acid sequence of the protein it codes for. Expressing the genetic code requires only one strand of DNA, the sense or template strand, which is copied into a complementary copy of messenger RNA (mRNA) which in turn directs protein synthesis at the ribosome (Figure 1.4); note the use of U instead of T in RNA.

Figure 1.4.



It has now been well established that eukaryotic DNA is stored in the nucleus of the cell, in a supercoiled form on pairs of chromosomes, containing around 10^6 - 10^8 nucleotide base pairs each, which store all the information necessary not only

to direct protein synthesis but to exercise control over the growth, division and differentiation of cells. This diversity of function is attributed to the flexibility of B-DNA which is able to interact with proteins and small molecules which control the mechanisms by which genes are switched on and off.

Despite the many advances in the four decades since B-DNA was first characterised many of its functions are still unknown. Only when the mechanisms for gene regulation / expression, cell differentiation and division are understood at the molecular level will it be possible to address such problems as mechanisms by which proto-oncogenes are converted into oncogenes, resulting in the development and uncontrolled growth of cancerous cells. A major development over the last decade, which has enabled such questions to be addressed in much more detail, has been the automation of DNA synthesis (Beaucage 1992). This powerful technique has facilitated work with larger and purer amounts of DNA of defined sequence, allowing techniques such as n.m.r. and X-ray crystallography to be used to investigate DNA structure in detail. Another advantage is that oligonucleotides containing non standard base pairs, eg. A.C and T.G, and modified bases, such as 5-Methyl dC and 5-Bromo dU, can be investigated and their effect on the helix geometry and oligonucleotide stability established. Oligonucleotides bound to drugs can be crystallised, (Leonard 1992a and Kennard 1991), thereby yielding information about DNA drug recognition and the very complex problem of drug design. X-ray crystallography has provided some ground rules for DNA structure by establishing the various helical forms of A, B and Z DNA as well as the characterisation of a variety of sequences containing non-complementary base pairings. This has given an insight into the diversity of inter base hydrogen bonding and the ability of the helix to accommodate structural variations.

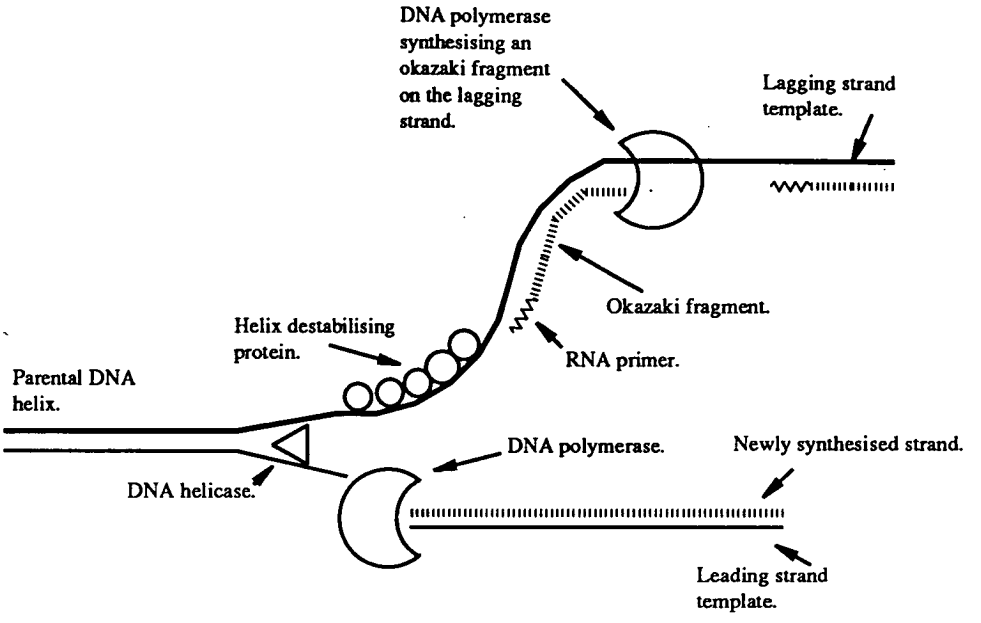
The importance of DNA in biology cannot be over emphasised and its role as a store of genetic information has been widely exploited in recent years, in techniques such as site directed mutagenesis (Taylor 1985), DNA fingerprinting (Joyce 1990) and in the use of antisense oligonucleotides to inhibit gene expression (Agrawal 1992). Developments in DNA sequencing, (Sanger, F. 1977), have led to the recent publication of the entire DNA sequence of yeast chromosome III (Oliver 1992) which contains 315 kilobases. Such work will undoubtedly contribute to our understanding of cellular processes which operate in both simple and complex organisms.

1.2. DNA structure and function.

Although many interactions between DNA, RNA and proteins have been characterised, many others are still not very well understood, yet are known to play a pivotal role in a variety of intracellular processes. A good example of the interplay between all three molecules is found in the series of reactions responsible for replication of the genetic code, (Figure 1.5), which is one of the most important events in nature. A huge variety of proteins are involved, eg. DNA helicase, helix destabilising proteins and DNA polymerase, for destabilising the double strand, stabilising the single strand and synthesising the daughter copy. RNA is also necessary as a primer on the lagging strand of the replication fork, possibly forming an RNA / DNA helix which can then act as a substrate for the DNA polymerase which only synthesises in the 5' to 3' direction. The entire process is also monitored by a variety of enzymes which ensure that the two daughter copies are synthesised complementary to the template strand. The fidelity of synthesis is such that there is only one wrong base in every 10^9 insertions (Fowler 1974).

Despite its structural and functional diversity, DNA is packaged in very compact

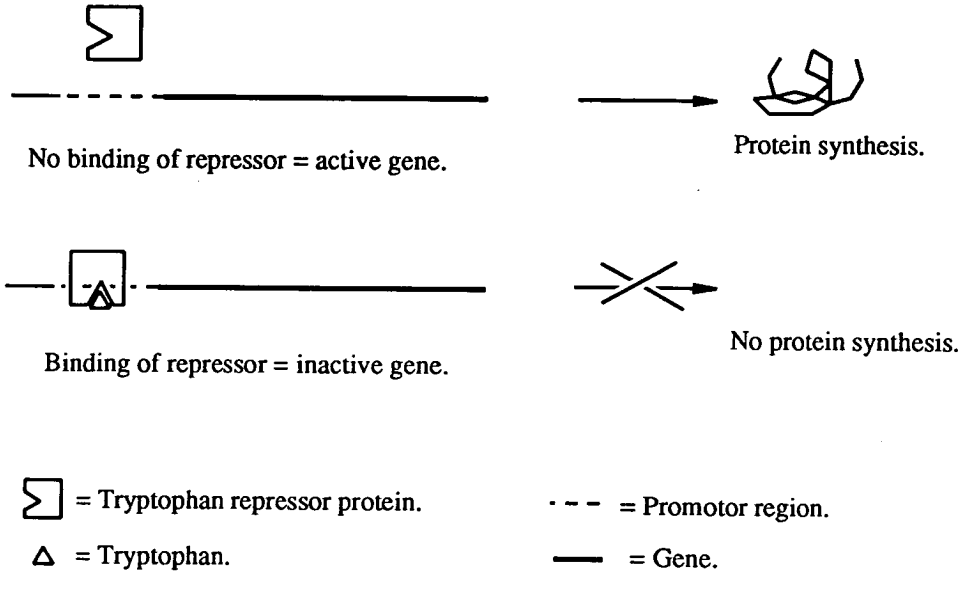
Figure 1.5.



supercoiled helices. This is the form in which DNA is stored in the nucleus of the cell as chromosomes, which consist of one long DNA molecule *ca.* 4 cm long compressed into fibres of *ca.* 25 nm wide by 700 μm long. This extremely dense structure of DNA is stabilised by proteins called histones which wrap the DNA up into very tight disks. When protein synthesis occurs an area of the chromosome is opened out and the double stranded B-DNA helix is exposed. DNA structure has profound implications for expression of genes and the control that DNA ultimately exerts over cellular processes as a whole, since in its active form DNA displays a wide range of structures with triple helices, cruciforms, hairpins and kinks all known to occur (Wells 1988), which may in some way be involved in the recognition processes involved in gene regulation. It has been well established that proteins recognise the double helical nature of DNA, often through a helix-turn-helix or a zinc finger motif (Alberts 1989), which have the ability to slide into the helical grooves under certain conditions, such as when small molecules or cations bind to the protein, and this leads to very tight binding to the DNA helix. Such processes are utilised by a class of protein molecules which activate or

repress gene expression by binding tightly to the promoter region of the gene. An example of repression, (Figure 1.6), shows how the genes coding for tryptophan synthesis proteins are inactivated when levels of tryptophan are high.

Figure 1.6.



The tryptophan binds to the repressor protein causing an allosteric change which enables it to bind tightly to the promoter region and prevent the binding of RNA polymerase to the single stranded DNA, thus preventing transcription and protein synthesis. When levels of tryptophan are low, competition for protein binding is diminished and the repressor loses its affinity for the DNA allowing RNA polymerase to bind and the gene to be activated. This process prevents the wasteful synthesis of the proteins, which are only required when levels of tryptophan are low and saves the energy resources of the cell by allowing amino acids to be recycled more efficiently.

In some way therefore, the protein must be able to distinguish between certain DNA sequences in order for it to bind specifically to the promoter region. This is certainly controlled to some extent by hydrogen bonding to functional groups on

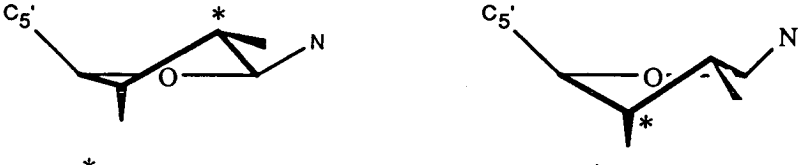
the base edges (Fliess 1988), but it is also possible that local tertiary structure may enhance this recognition. DNA is capable of forming very subtle tertiary structures which are a direct result of inter and intra strand base / base interactions finely adjusting the helix geometry in an effort to resolve the local steric clashes of functional groups created by certain base sequences. The forces of hydrogen bonding and base stacking are the horizontal and vertical forces, respectively, between the bases within the duplex and are the major contributors to geometry. Thus it can be said that DNA and RNA are not conformationaly homogeneous molecules since the observed structure is dependent upon the base sequence.

1.2.1. Hydrogen bonding.

Inter-strand hydrogen bonding holds the two strands of the duplex together. The enormous number of hydrogen bonds present, in even a small length of DNA, ensures that globally the two strands are held very tightly together and there is no possibility of them coming apart. The very fact that hydrogen bonding is by nature a fairly weak interaction, varying from 3-6 kcal/mol, does however mean that short sections of DNA can be easily opened into single strands allowing the information to be efficiently accessed without affecting other parts of the helix. The beautiful symmetry of B-DNA (Figure 1.3), is a direct result of the complementary hydrogen bonding across the helix axis, which holds the duplex together giving it a psuedo two fold axis of symmetry. With two antiparallel strands the dyad symmetry is restricted to the C1'-N (glycosidic) bond and is not applicable to the actual bases, hence it is called a pseudodyad. Such a high degree of pseudo symmetry is only possible if, as is usually the case, the strands can form the naturally pseudo symmetric Watson / Crick pairs, (Figure 1.2.). The hydrophobic planar bases are stacked up the core of the helix, to the exclusion of water, whereas the negatively charged hydrophilic phosphodiester-sugar

backbone winds around the outside of the helix where it can interact with water and cations which are always present *in-vivo*. This interaction with water has a pronounced effect on the overall structure of the helix since at high, (90%), humidity it forms a B helix with each phosphate groups hydrated by individual water molecules. At lower, (70%), humidity the helix exists in the more contracted A form whereby the phosphates are too close to be complexed by individual water molecules and are stabilised by bridges of water molecules between neighbouring phosphate groups. The separation between the phosphate atoms, on neighbouring bases of the same strand, is reduced from 700 pm to 550 pm, and is accompanied by an alteration of the puckering of the sugar moiety from C2'-*endo* to C3'-*exo* (Figure 1.7). It is of interest to point out that to date double stranded RNA has only been found to exist as an A-form helix, which is thought to be a result of the inability of ribosides to adopt a C2'-endo pucker

Figure 1.7.



*C2' - endo. *C3' - exo.

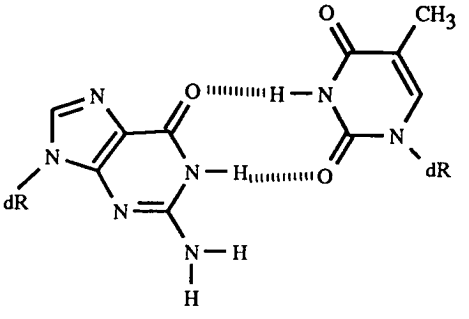
because of the 2'-hydroxyl functionality. Many short sequences of DNA have been characterised by X-ray crystallography in A, B and Z-DNA forms, for review see (Kennard 1991 and Wing 1980).

It would appear however, from this crystallographic work, in conjunction with NMR, that the structural diversity of DNA is primarily due to the flexibility of the B-helix and its ability to accommodate minor perturbations, whereas the more rigid A-helix is found to be less flexible (Saenger 1984). There is also a well characterised spine of hydration in the minor groove of the AATT region of the B form

d[CGCGAATTCGCG] dodecamer (Kopka 1983 and Drew 1981). It is thought to play a significant role in stabilising the B conformation by preventing it from opening out into a shallow groove which is more characteristic of the A form helix. It is particularly important in such dA.T rich regions, which are known to depart from the usual Watson / Crick geometry by inducing bending along the helical axis (Nelson 1987).

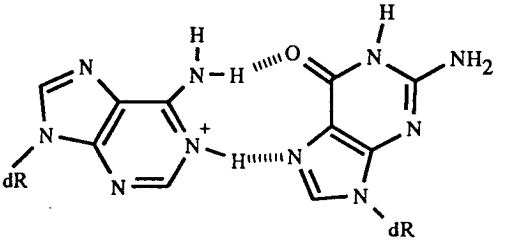
In addition to the four Watson / Crick base pairs, many other pairings are possible which can have dyad, pseudodyad or no symmetry but which destabilise double stranded DNA. Such non standard base pairs have been incorporated into otherwise self complementary sequences to give double helical oligonucleotides with two mispairs, example d[GGGTGCCC]. which is a self complementary oligonucleotide which forms an A-DNA duplex containing two T.G mispairs in the centre (Rabinovich 1988). The high resolution structure of this, and other related sequences, have been determined by X-ray crystallographic methods and two mismatched base pairs, characterised from such studies, are illustrated below, Figure 1.8a and 1.8b.

Figure 1.8a.



dG : T

1.8b.



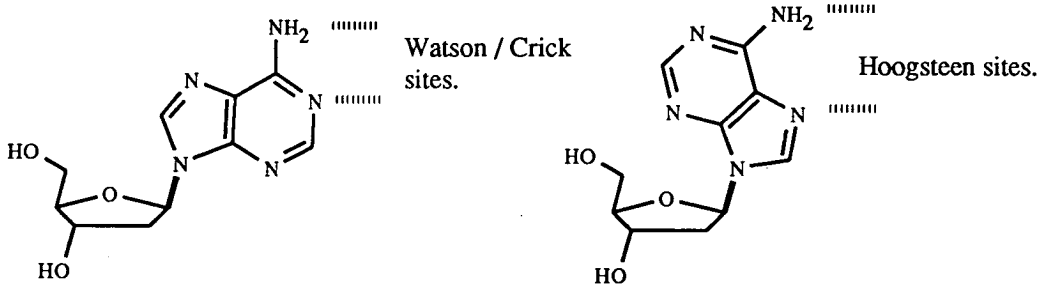
Protonated dA (*anti*) : dG (*syn*)

The dG.T mispair, (Figure 1.8a.), is held together by two hydrogen bonds. In order to accommodate them however, it is necessary for the thymine base to move

into the major groove thus forming what is termed a wobble base pair. As a result, the pseudodyad symmetry of the base pair is lost, (compare with the dG.dC and dA.dT pairings in Figure 1.2 which maintain pseudo-symmetric centres). In contrast the dA.dG mispair, (Figure 1.8b.), maintains its pseudo symmetric centre, also displaying conformational variability as a function of the pH of the solution (Gao 1988). Base stacking also plays a very important part in the geometry and structure of this mismatch, and this effect is discussed in the next section, 1.2.2., (Brown 1989 and Brown 1986).

The variations observed for this mismatch are in part due to the *syn / anti* conformational change that is possible around the C1'-N glycosidic bond, (bold bond in Figure 1.9), and in purine bases this allows alternative hydrogen bonding N and O atoms to be presented to the opposing base.

Figure 1.9.



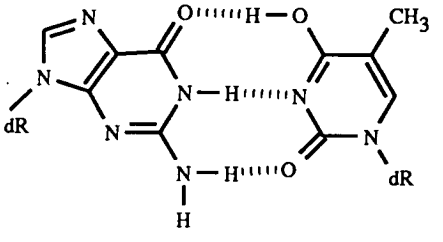
dA (*anti*). dA (*syn*).

This *syn / anti* rotation permits a variety of dG.dA pairings, of which four have been characterised (Ying 1991, Brown 1989 and Brown 1986). One of them, dG(*syn*).dA(*anti*), (Figure 1.8b), is held together by two hydrogen bonds, one of which is only made possible by protonation of the dA-N(1) atom, and rotation of the dG base into the *syn* conformation. This base pair utilises the N(7) and O^6 hydrogen bonding sites of dG which are complementary to the N^6 and N(1) atoms

of protonated dA. The separation between the C1' atoms of the dG and dA deoxyribose sugars is approx. 11 *A*. This is close to the value for a Watson / Crick base pair (10.1 to 10.8 *A*) and so ensures that the helical diameter is maintained without disruption of the phosphodiester backbone. It is however a thermodynamically unstable helix, T_m = 293°K (pH 7.75) and 311°K (pH 5), and crystallisation of this form of mismatch is only possible below physiological pH where the base pair is able to form the hydrogen bond involving the protonated N(1) atom of dA which is necessary for stabilisation of the helix (Brown 1989). The conformational variability of the dG.dA mismatch illustrates the variety of methods employed by DNA to accommodate unusual structures into the local helical geometry by fully utilising the limited number of functionalities in the heterocyclic bases capable of participating in hydrogen bonds.

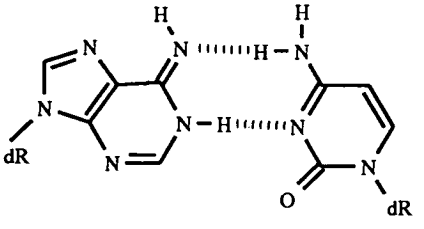
Such mispairings, as illustrated in Figure1·8, provide a mechanism for DNA mutation. An alternative mechanism, *via* formation of base pairs which contain one of the bases as a minor tautomer is shown below, Figure 1.10.

Figure 1.10a.



dG . T(enol)

1.10b.



dA(imine) . dC

In this situation a dG can pair with a T base in the enol tautomeric form to form three stable hydrogen bonds and a dA imine tautomer can pair with dC to form two stable hydrogen bonds. Both provide a mechanisim for the incorporation of a different base sequence into the daughter strand at the next round of replication.

The alternative form, dG(enol).T and dA.dC(imine) will have the same effect.

The almost exclusive formation of Watson / Crick base pairs in DNA is extremely important during replication and it is still unclear how enzymes are able to detect mismatches so efficiently in spite of the tremendous variety of mismatch structures that can form. Variable hydrogen bonding patterns are far more extensively utilised in RNA molecules than in DNA. RNA displays extreme conformational flexibility due to the presence of the 2' hydroxyl group which enhance its hydrogen bonding capabilities. This enables it to fold into a variety of very complex single stranded tertiary structures and consequently RNA has many more diverse functions in biology than DNA. It is utilised in ribosomes, which are large RNA molecules containing three RNA binding sites and which synthesise proteins of sequence defined by the bound mRNA strand (Alberts 1989). A further demonstration of the diversity of RNA is demonstrated by ribozymes which are highly specialised RNA molecules having a well defined tertiary structure and are capable of catalysing intra-strand phosphodiester bond cleavage (Cech 1986).

1.2.2 Base stacking.

Hydrogen bonding controls the recognition that enables two complementary strands to form a double helical structure. However, after initial nucleation of *ca.* three base pairs the vertical base stacking interactions further stabilise the complex and promote full duplex formation (Porschke 1977). These interactions are extremely important in determining the secondary structure, which therefore depends upon the the DNA sequence. The situation regarding mRNA and tRNA is complex due to the extensive use of hydrogen bonding to maintain structural integrity. In certain DNA sequences, such as inverted repeats or G rich regions, intra strand hydrogen bonding and base stacking can occur to give the fairly well

defined structures found in telomeres and hairpins (Kang 1992 and Howard 1991). The subtle perturbations imparted by stacking within the duplex, which are very dependent upon the base sequence involved, ensure that the planar base moieties are stacked in such a way as to keep interaction with water and steric clashing of functional groups to a minimum, thus lowering the free energy of the system. Stacking forces depend upon; dipole / dipole, dipole / induced dipole interactions and London dispersion forces (Atkins, P.W., 1985), as well as interactions between the π-electron systems and hydrophobic effects (Hanlon, S. 1966 and Herskovits 1963). The sequence dependence of base stacking effects can be seen quite clearly from the stacking energies of all the ten possible base pair dimers, which have been calculated *in vacuo* (Ornstein 1978). These results clearly demonstrate the importance of sequence on the energy of stacking, and the following general rule applies:

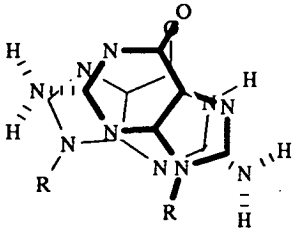
5'-Pyrimidine-Purine-3' more stable than **5'-Purine-Pyrimidine-3'**

This is typified by the 5'-CG-3' dimer which is 5 kcal/mole lower in energy than the 5'-GC-3' dimer, and is a direct result of the lack of overlap found between bases when the pyrimidine is found on the 3'- side of a purine base (Arnott 1974). This effect is only valid for the B-DNA form since base pair tilting, propeller twisting and helical rise per residue vary for each form and have a major effect on the base stacking potential (Dickerson 1983). For example pyrimidines situated 3' to purines give rise to functional group clashes in the minor groove which can only be alleviated by local opening up of the helix by bending of the DNA towards the major groove. Such bending of the helix has been shown to be recognised by certain restriction endonucleases which preferentially cut at TpA or CpG steps. The sequence recognised by*Eco*RV, for example, is GA<u>TA</u>TC, and cutting occurs between the underlined bases. It has been demonstrated however

that *Eco*RV binds to all sequences with equal affinity (Taylor 1991) but will cut only at the TpA step because of preferential binding of Mg^{2+} to this DNA / enzyme complex. The catalytic behaviour of this enzyme has also been studied using modified nucleotides in the vicinity of the TpA region, and the results have shown the importance of cytidine and adenine amino groups for enzymic activity (Fliess 1988). It is therefore conceivable that slight bending of the helix, coupled with recognition of the flanking bases, induces a conformational change in the enzyme / DNA complex which allows binding of Mg^{2+} and thus catalysis of the reaction. There is almost no affinity for Mg^{2+} binding at other sites (Taylor 1989).

It has also been observed that stacking plays a very important part in the crystal packing of free nucleosides, with the stacking pattern varying widely depending upon the nature of the parent molecule. Stacking interactions observed for guanosine (Bugg 1968) are shown (Figure 1.11a), and inosine, which lacks the N^2 amino group, follows the same pattern. On the other hand the crystal structure of 8-bromo-2'-deoxyguanosine shows a completely different stacking pattern to that of the parent molecule, (Figure 1.11b), (Bugg 1969).

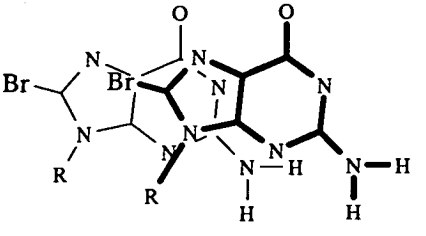
Figure 1.11a.



R = Ribose.

Guanosine (Inosine).

1.11b.



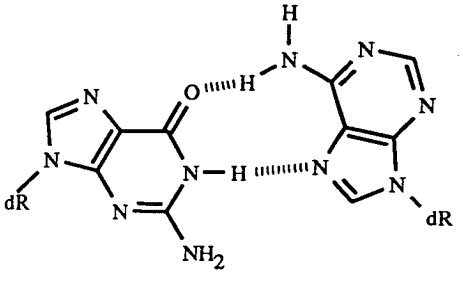
8-Bromoguanosine.

The introduction of the very electronegative bromine atom creates a polarisable bond which is able to interact with the π-electron system through an inductive

effect (Wallwork 1961) and as a result the base stacking is completely altered.

The effect of the stacking environment on the structure of individual base pairs can be highlighted by the differences in structure of dG.dA mispairs found in the following pair of related dodecamers: A) d[CGCGAATTAGCG] and B) d[CGCAAATTGGCG]. Sequence A gives rise to a dG(*anti*).dA(*syn*) mispair whilst B results in formation of a dA(*anti*).dG(*syn*) mispair. The flexibility of this mismatch has been demonstrated through NMR experiments (Gao 1988), and it is clear from this and the crystal structure results, that base stacking effects, along with pH variation, influence the conformation. This demonstrates the ability of flanking bases to exert pressure on the mismatch forcing the bases to adopt the conformation most able to enhance the duplex stability. In the dG(*anti*).dA(*syn*) mismatch, (Figure 1.12), the dG involved is part of a CpG step which is known to be the most stable stacking step in B-DNA (Ornstein 1978), and is almost certainly partly due to the fact that cytosine and guanine have dipole moments of equal magnitude but opposite polarity (Brown 1989).

Figure 1.12.



dG (*anti*) . dA (*syn*).

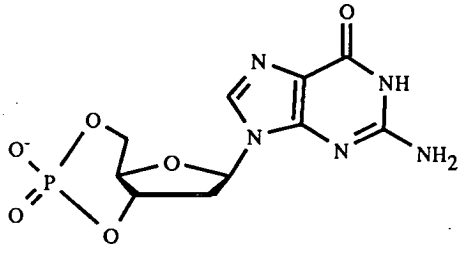
In sequence B however, this stable stacking is not present because the dG is now part of a GpG step and so unfavourable dipole interactions are best relieved by rotation of the guanine into the *syn* conformation which will more favourably align their static dipole moments, to yield the protonated base pair shown in

Figure 1.8b. This alignment is illustrated by the crystal structure of guanine which stacks as in Figure 1.11a.

1.3 DNA modifications.

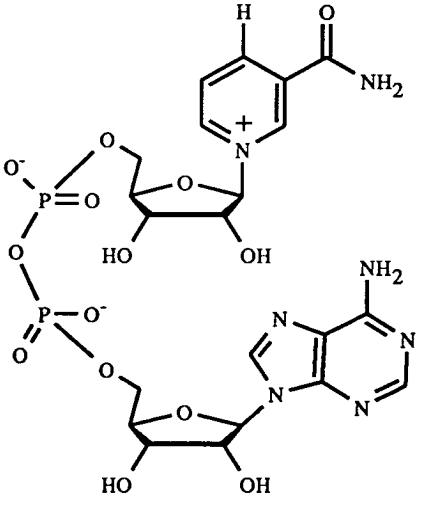
DNA has a large number of functional groups, ie:- keto, amine and hetero atoms, capable of reacting with a wide variety of chemicals. Although all exocyclic groups and hetero atoms, with the exception of the nitrogen attached to the sugar, are available for reaction (Blackburn 1986), modification of heterocyclic bases is much more common than that of the sugar or phosphate group since it is thought that disruption of the very stable phosphodiester backbone may inhibit the formation of double helical DNA, and hence such alterations are easily recognised and repaired. In addition, sugar / backbone modifications often lead to cleavage of the strands, a lesion which is also very easily recognised and repaired. The most common chemical modifications involve nucleophilic attack on electrophilic alkylating agents by heterocyclic bases. These are the most extensively found potentially mutagenic modifications.

Figure 1.13a.



cGMP.

1.13b.



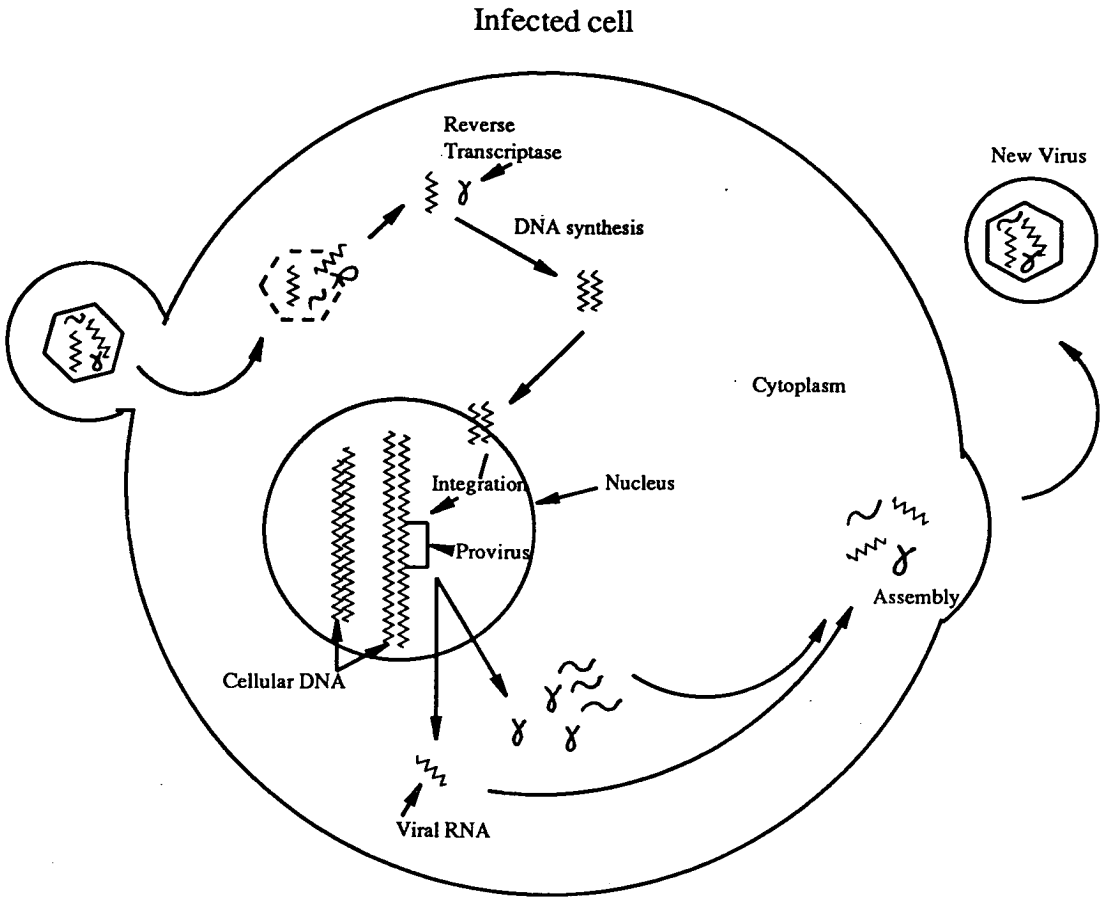
NAD^+

Modifications of this nature also occur naturally in tRNA (McCloskey 1977) where enzymes selectively modify the RNA nucleotides in order to convert the tRNA into a functional molecule. Other derivatives such as 5',3' cGMP and NADH are utilised as secondary hormonal messengers and enzyme cofactors respectively. (Figure 1.13).

1.3.1 Phosphate and sugar modifications.

Research into the synthesis of novel nucleosides has expanded greatly in the last decade, largely due to the isolation of the Human Immunodeficiency Virus (HIV) in 1982, and hundreds of nucleoside analogues have been prepared and evaluated as *anti*-HIV agents. HIV is the virus responsible for AIDS (Montagnier 1988), and it belongs to a class of viruses known as retroviruses because its genetic information is stored as RNA as opposed to DNA. In order for the virus to complete its lifecycle (Figure 1.14), ie. to create more viral particles, it is first of all necessary to convert its information into double stranded DNA. After fusion of the virus with the host cell, viral DNA is synthesised by the enzyme reverse transcriptase which uses both viral tRNA, as a primer, and viral RNA as a template for the synthesis of a DNA / RNA hybrid. Subsequently a DNA / DNA duplex is synthesised which is then integrated into the host genome where it can remain dormant for many years before transcribing its information and producing new viral particles which results in lysis of the host cell. This very subtle lifecycle of HIV explains the difficulty in controlling its spread since *anti*-HIV drugs must be harmless to uninfected cells. The enzyme reverse transcriptase, which is injected into the host cell along with the RNA inside the virus particle, is a good target for this inhibition strategy because it is unique to the infected cells. It is not a particularly selective polymerase (Preston 1988), as abnormal nucleotide triphosphate analogues can be incorporated into the viral genome far more readily

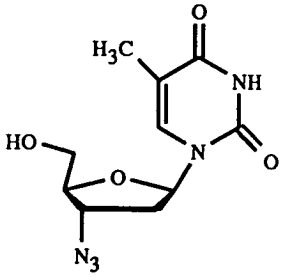
Figure 1.14.



than into host DNA which is synthesised with the aid of cellular polymerases. Currently, the most potent inhibitor of reverse transcriptase is 3'-azidothymidine (Figure 1.15a), a thymine analogue which when attached to a growing DNA strand cannot support chain elongation due to the absence of the 3'-hydroxyl group, thus polymerisation ceases and only short fragments of DNA are produced. 3'-Azidothimidine is incorporated into viral DNA 100 times faster than host DNA which is synthesised by cellular polymerases (Yarchoan 1988). Acyclovir, (Figure 1.15b), is an inhibitor of Herpes Simplex Virus (HSV), and in this case selectivity to virally infected cells arises from the fact that it can act as a substrate for HSV thymidine kinase but not for cellular thymidine kinase. Thymidine kinase is the

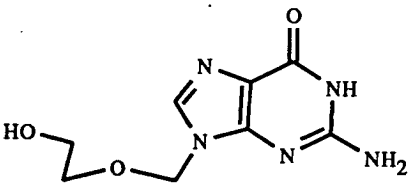
enzyme responsible for the phosphorylation of nucleosides to produce their monophosphate derivatives (Elion 1989).

Figure 1.15a.



Azidothymidine.

1.15b.



Acyclovir.

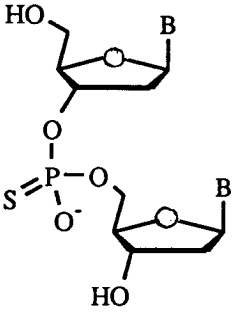
In both AZT and acyclovir the active compounds are the nucleotide triphosphates which are the substrates for the polymerase enzymes and can be incorporated into the growing strand. Although extremely potent, these drugs do not produce cures and their quick turnover *in-vivo* means that they must be administered daily. Thus treatment is long term and very expensive.

An alternative antiviral strategy is known as the antisense approach which involves the synthesis of an oligonucleotide >17 bases long which has a complementary sequence to a coding region of the viral genome. This subject has been extremely well reviewed recently (Uhlmann 1990). The antisense strategy works by the binding of an oligonucleotide to mRNA to prevent protein synthesis at the ribosome, or by binding of an oligonucleotide to the genomic DNA by triple helix formation, thus preventing synthesis of mRNA (Riordan 1991). Although both methods of inhibition are perfectly acceptable models for arresting the viral life cycle, progress in this area has been fraught with difficulty due to the inability of the oligonucleotides to penetrate the cell membrane as well as their lack of resistance to degradation by DNase enzymes both inside and outside the cell. Many chemical modifications have been made to antisense oligonucleotides to combat these problems, mainly the incorporation of large lipophilic molecules

such as vitamin E and cholesterol (Will 1992 and Stein 1991) in an effort to improve transmembrane transport, and the incorporation of phosphorothioate groups to prevent cellular degradation, (Figure 1.16), (Ott 1987). Another interesting approach has been the attachment of intercalating molecules such as psoralen and acridine to oligonucleotides (Englisch 1991), such that the site of intercalation is controlled by the oligonucleotide binding in a triple helix fashion to genomic DNA.

Figure 1.16.

Phosphorothioate linkage.

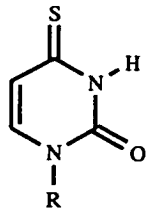


Other strategies have been devised in an attempt to couple improved lipophilicity with nuclease resistance. Oligonucleotides have been synthesised containing a neutral backbone (Matteucci 1990 and Sood 1990). A combination of improved lipophilicity and nuclease resistance has been achieved, although many questions still remain as to the mode of action of these analogues (Agrawal 1992). Such molecules do however have very good potential as antisense agents and this has been demonstrated by their efficacy when microinjected into infected cells.

1.3.2 Heterocyclic base modifications.

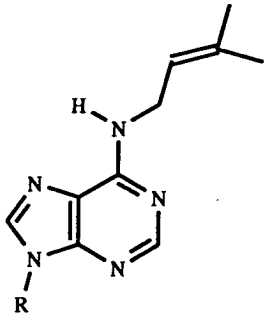
Transfer RNA is an extremely rich source of unusual ribosides (Figure 1.17). The most common modification involves the conversion of a keto group into a thiol and alkylation of exocyclic amines (Saenger 1984).

Figure 1.17.

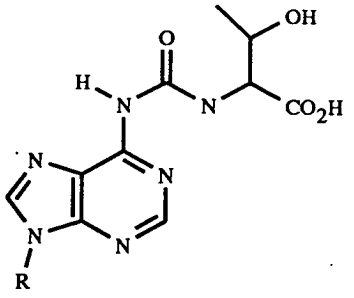


4-Thiouridine.

R = ribose.



N^6-isopentenyladenosine.



N-[N-(9-β-d-ribofuranosyl-purine-6-yl)carbonyl]threonine.

Modified ribosides are present at particular sites throughout tRNA molecules but are clustered predominantly around the anticodon domain. They are necessary for enhancing the codon / anticodon interactions with the modification being dependent upon the first nucleotide of the codon it recognises in the mRNA sequence. In particular, the flanking base to the codons exhibit the greatest variety of modifications, for example codons starting with a U almost always contain hydrophobic modifications, eg. N^6-isopentenyladenosine, whereas tRNA recognising codons starting on A contain hydrophilic modifications such as N-[N-(9-β-d-ribofuranosylpurin-6-yl)carbamoyl]-threonine. Modified bases at other locations may also promote the formation of very precise tertiary structure by blocking and enhancing hydrogen bonding sites that are important in the formation of base triplets (Altman). In addition modified bases inhibit the degradation of tRNA by RNase enzymes.

Synthetically prepared nucleoside derivatives containing reactive functionalities, eg. sulphur or azido, can be incorporated at specific points in a sequence and are particularly effective at promoting intra strand and DNA / protein crosslinking (Glick 1991 and Meffert 1990). Meffert synthesised 8-azidoadenine and incorporated it into a sequence containing the tetracycline repressor binding

domain. Binding of the repressor followed by long wave irradiation yields a protein-DNA crosslink which is easily visualised on gel electrophoresis. It has also been used in its triphosphate form to probe the deoxynucleotide binding domain of the enzyme DNA polymerase (Joyce 1987). This methodology therefore has the potential to provide information about protein binding domains, tolerance of DNA modifications on protein binding and also kinetics of gene repression / derepression. On the subject of the evolutionary development of DNA and RNA as replication systems, much work is currently underway on the introduction of novel, but complementary base pairs, which can extend the genetic alphabet beyond the four Watson / Crick base pairings (Piccirilli 1990).

A more detrimental aspect of nucleoside modification, discussed in the next section, is the formation of derivatives which contain an altered hydrogen bonding pattern with respect to the parent compound and thereby lead to the formation of mutagenic lesions. The synthesis of such nucleosides is extremely important and the effect they have on double helical DNA must be investigated thoroughly so that their mutagenic potential can be accurately assessed. The crystal structure, (Leonard 1992b), and NMR analysis, (Patel 1986), of such sequences gives us an insight into the altered hydrogen bonding, base stacking and base orientations which cause mispairing or reading frame errors and lead to mutagenic lesions in DNA.

1.4. Alteration of the genetic code.

As mentioned briefly in section 1.2, the outstanding accuracy of DNA replication is such that only one base in every 10^9 insertions results in a mutation in the daughter helix. Mispairings, such as dA.dC or T.dG pairs, do not normally persist all the way through the replication cycle and give rise to mature DNA because a variety of mechanisms have evolved in cells to counteract this problem. Initially

the base selection is achieved through the parent strand acting as a template for incoming nucleoside triphosphates, which, when they match the complementary hydrogen bonding pattern are attached to the 3' end of the growing chain. This is however a very crude method of selection since it is based solely on the energy difference between a complementary and a non-complementary base pair during DNA synthesis, which is around 3kcal/mole ie. the equivalent of one hydrogen bond (Loeb 1981). Therefore, the presence of minor tautomers, which have a frequency of *ca.* 10^{-5}, and base mispairings can cause the normal rules of Watson / Crick complementarity to break down, resulting in an initial misinsertion frequency of 10^{-2} to 10^{-4}. The most important controlling influence on overall fidelity is the DNA polymerase protein itself which has both a 3'-5' and 5'-3' exonuclease domain with the ability to recognise base mispairs, cut them out and insert the correct base to lower the error rate to an acceptable level.

Since genes are made up of non-coding lengths called introns and coding lengths called exons, depending upon where a mutation occurs it may or may not directly affect the integrity of the cell. All mutations, however, regardless of their effect can be classed as either transition mutations (arising from the pairing of a purine with the wrong pyrimidine) or transversion mutations (arising from the pairing of two purines). The twelve possible base pair conversions are shown, (Figure 1.18).

Figure 1.18.

Transversion mutation. Transition mutation.

When such a mutation occurs in an exon region it is known as a point mutation and has the potential of altering the amino acid sequence due to a change in the

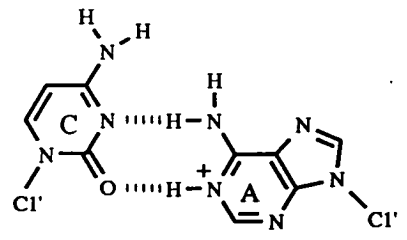
triplet code. When it occurs in an intron region that is in the vicinity of a gene, eg. the promotor region, then it is termed a regulatory mutation since the introduction of a different base in such an area may have the effect of determining whether the gene is active or not. An interesting aspect of DNA mutations has been their use in establishing a timescale for the divergence of various species (Alberts 1989). In all species, mutations will build up at a steady rate, thus when the relationship between frequency of mutation and time is established an estimate can be made of the lineage of the species. Within humans the relatively fast rate of mutation in mitchondrial DNA has been used in an attempt to establish the timescale of divergence of modern man and Neanderthal man into the present day races (Vigilant 1991).

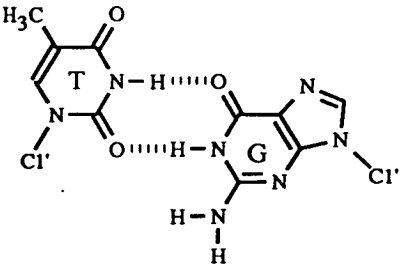
Persistence of mutagenic base pairs is closely related to their ability to mimic normal Watson / Crick base pairs in terms of their overall geometry in relation to the helix. The three mispairs illustrated in Figure 1.19 are all repaired to varying degrees. The dG.T and dA.dC wobble pairings, whose geometry least closely resembles a true Watson / Crick pair (Fersht 1982), are repaired more frequently than the dA(*syn*).dG pairing (repair ratio of 15:5:1) which is most likely to escape detection by the proofreading enzymes.

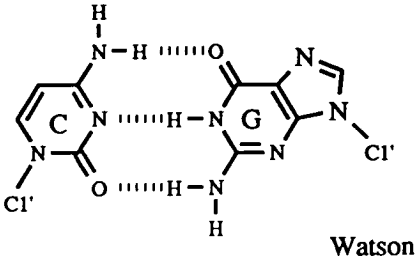
In a crystal structure of the dA.dC mismatch (Hunter 1986) it is noticeable that the mismatch has a major impact on the stacking properties of the adjacent dC(3) . dG(21) bases which are altered from the norm, and it is entirely possible that if this event were to occur *in-vivo* then misreading could also occur at those positions.

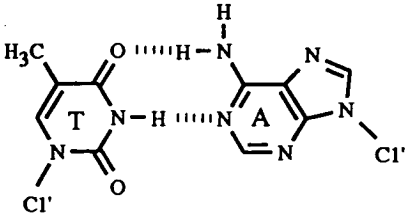
Although mutations slowly accumulate from one generation to the next, there is a certain amount of self regulation because the pressure to mutate a particular base is very dependent upon what region of the enzyme for which it is coding. The active site

Figure 1.19.

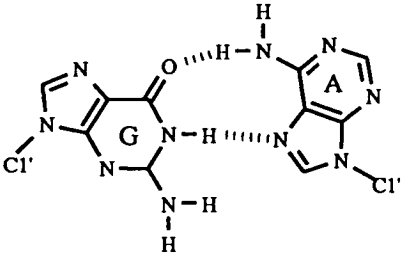








Watson / Crick base pairings.



Geometry of base mispairs in comparison to the Watson/Crick pairs.

of an enzyme will show almost no tolerance of mutation whatsoever since any change would most probably result in the enzyme working less efficiently. For the same reason, residues involved in protein folding will show similar resistance. Amino acids in less structurally important regions of the enzyme, such as side or linking chains, will however be much more susceptible to change as any changes may not directly affect the integrity of the enzyme. Thus the protein sequence, of for example haemoglobin, from a variety of animals shows areas of strong conservation in the structurally important regions but little or no conservation in structurally inactive portions. This disparity is often used to establish time scales of species divergence. It is important to point out however that in principle all

areas of a protein, eg. an enzyme, are subject to the same frequency of mutation, it is just that when they occur at substrate binding sites they will usually be detrimental to the organism and the mutation will not be maintained in the gene pool. This "protein survival of the fittest" therefore ensures that detrimental mutations are cast aside whilst beneficial ones are maintained due to optimisation of the enzyme's catalytic activity.

An excellent example of how environmental pressure selects in favour of mutations is found in inhabitants of West Africa, where malaria is prevalent. It is therefore beneficial to be able to withstand the mosquito borne attacks of the parasite responsible for malarial infection. Consequently, a dG.dC to T.dA transversion mutation in the gene coding for human haemoglobin, has resulted in the conversion of a single d(GGA) triplet coding for glutamic acid, to a d(GTA) triplet which codes for valine. This incorporation of valine instead of glutamic acid results in the blood corpuscles becoming less saucer shaped than normal, making them less susceptible to invasion by the parasite (Smith 1975). This mutation also has a detrimental effect since this "sickling" of the corpuscles can block blood vessels and this in turn may lead to anaemia and even death. It is important to point out however, that although this mutation is both beneficial and detrimental, it has survived in the population because it has given them some resistance to malaria, which causes far more deaths than sickling.

Mutations causing gene alterations, such as described above, give rise to problems such as Cystic fibrosis (Cheng 1990 and Gregory 1991) and Tay-Sachs disorder (Lau 1989) which are illnesses resulting from the inheritance of a defective gene. The resultant protein is structurally different and this can prevent full maturation leading to disruption of the usual protein trafficking mechanisms.

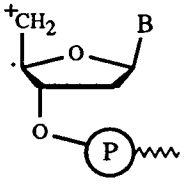
In summary, there is a low background level of mutation which gives rise to a large gene pool within a species and enables it to respond quickly to changes in environmental factors. If replication was 100% faithful then no evolution would have been possible due to the absence of a mechanism for improvements through preferential selection. Despite these errors in copying, inheritance of genetic material from one generation to the next proceeds very faithfully with mutations occurring fairly infrequently. The study of genetic inheritance is therefore carried out on species such as *Drosophila melanogaster* which have short life spans allowing many generations to be studied over a period of a few months (Garcia-Belido 1979). Such studies yield information that allows genes to be accurately mapped out, ultimately increasing our knowledge of proto-oncogenes and hereditary defects.

1.5. Mutation through DNA damage.

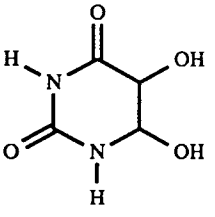
Base mispairs which escape detection by polymerase and repair enzymes account for a very small number of mutational lesions and merely maintain a certain background level. A much more serious threat to the genetic integrity of DNA arises through chemical or physical damage, particularly during processes such as replication, transcription and recombination where the genome is partially denatured from proteins in its densely supercoiled form into the more accessible protein free double stranded form. As a result of the constant presence of mutagenic agents in the environment, which affect not only DNA but many other biological molecules, a variety of detoxifying compounds and repair enzymes have evolved to prevent and repair damage caused by this constant daily exposure. Acute attacks from radiation exposure and ingestion of toxic chemicals can however completely overload these defence systems leading to cellular death, through damage to proteins, DNA and consequently tissues. The interactions of

physical and chemical agents with DNA are particularly diverse owing to the large number of sites throughout the DNA molecule which can be easily attacked. Both classes of carcinogen can however ultimately react by inducing some form of structural change in the bases such that a mutagenic lesion is formed which has lost the base pairing complementarity of the parent compound. Physical damage due to for example, ionising radiation, can lead to the formation of very reactive oxygen radical species which can damage DNA through strand scission (Dizdoraglu 1975) as well as causing a variety of nucleotide modifications (Aruoma 1991), (Figure 1.20).

Figure 1.20.



Strand scission.

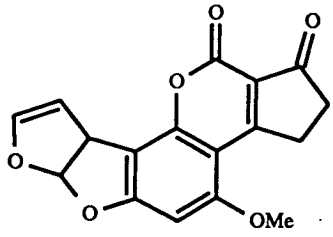


(cis/trans)-Thymine glycol.

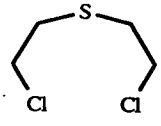
Electrophilic alkylating agents constitute the major group of chemical carcinogens. They are capable of reacting at a multitude of nucleophilic sites in DNA (Blackburn 1986). These agents are extremely diverse, from intercalative polyaromatic compounds such as benzo-(a)-pyrene and anthracycline drugs to aflatoxins and sulphur / nitrogen mustards, (Figure 1.21).

The latter compounds are bifunctional electophiles which exploit the high nucleophilicity of guanine N(7) and severely damage DNA by crosslinking the two strands. The principle of chemical crosslinking of opposite strands is utilised by the *anti* cancer drug cis-Platin (Lippard 1979).

Figure 1.21.



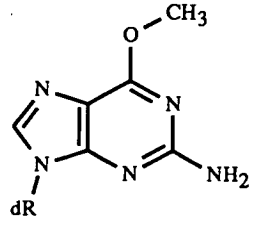
Aflatoxin B1.



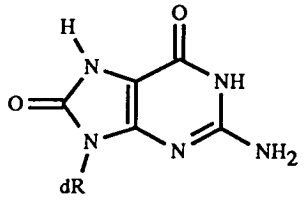
Sulphur mustard.

Two compounds which form mutagenic lesions are O^6-methyl-2'-deoxyguanosine (O^6MedG) which is the result of nucleophilic attack on electrophilic alkylating compounds and 7-hydro-8-oxo-2'-deoxyguanosine (8OxodG), which is a result of radiation or oxidative damage to DNA, (Figure 1.22).

Figure 1.22.



O^6-Methyl-2'-deoxyguanosine.



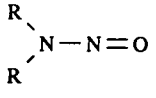
8-Oxo-7-hydro-2'-deoxyguanosine.

Both are known to be mutagenic, O^6MedG specifically induces G.C to A.T transition mutations (Snow 1984 and Abbott 1979), whilst 8OxodG directs the insertion of dA or dC depending upon the nature of the polymerase (Shibutani 1991). It is therefore very important that the mechanism for misinsertion by these mutagenic bases is elucidated so that a greater understanding of the onset of cancers can be established. Only through structural analysis at the base pair level can their effects on duplex DNA be investigated.

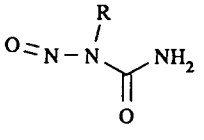
1.5.1. Alkylation of Guanine O^6.

Electrophilic alkylating agents are classed in two groups, those which can directly react with DNA, (primary carcinogens), and those which are metabolised into reactive compounds, (secondary carcinogens). Compounds responsible for a higher proportion of the mutagenic O^6- alkylation almost exclusively come from the latter group and are typically N-nitrosoureas and dialkylnitrosamines which are metabolised into very reactive alkyl diazonium ions, (Figure 1.23a). Such compounds are termed 'hard' electrophiles and the ratio is overwhelmingly in favour of N-alkylation. However there is sufficient S_N1 character for the fraction of O^6- alkylation of dG to N(7) alkylation of dG to be *c.a.* 0.1, with methyl-nitrosourea as the alkylating agent (Singer 1982a). In contrast, primary carcinogens such as dialkyl sulphates and alkyl-alkane sulphonates, (Figure 1.23b), are much softer electrophiles and react in typical S_N2 manner such that the fraction of O^6- alkylation to N(7) alkylation is lowered to *ca.* 0.003, with dimethyl-sulphate as alkylating agent (Singer 1982a).

Figure 1.23a.

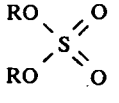


Dialkylnitrosamines.

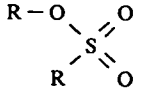


N-Nitrosoureas.

1.23b.



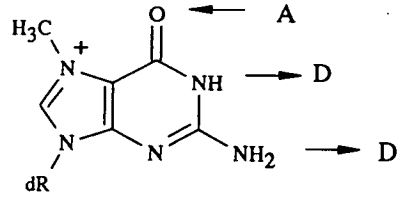
Dialkyl sulphates.



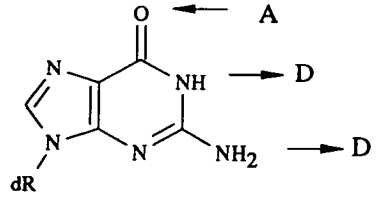
Alkyl alkane sulphonates.

Of the five sites available to alkylation in deoxyguanosine, N(1), N^2, N(3), O^6 and N(7), the O^6 position is the one for which substitution produces the major mutagenic lesions (Singer 1982a), Such modified bases lead to the insertion of thymine rather than cytidine in the opposite strand during DNA replication. A mutagenic lesion of this nature is known in one particular example to initiate the onset of a malignant change in cell character through activation of the H-ras 1 transferring oncogene (Sukumar 1983). Although N(7) methylation occurs far more frequently than O^6 methylation (Singer 1982b) it does not generate a mutagenic lesion, probably because there is no change in the hydrogen bonding complementarity of the modified nucleoside to deoxycytidine (Nagata 1988). N(7)-alkylation is a particularly easy lesion to repair since it leads to strand scission and the resultant abasic site can be easily repaired by AP, (apurinic / apyrimidinic), endonuclease activity (Friedberg 1985). Figure 1.24 shows how N(7) methylation of dG does not alter its ability to hydrogen bond to dC since the Watson / Crick sites remain unaffected by quaternisation at the N(7) position.

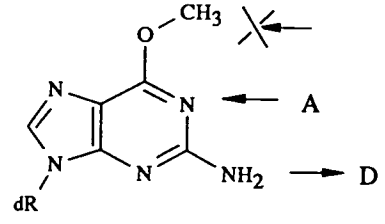
Figure 1.24.



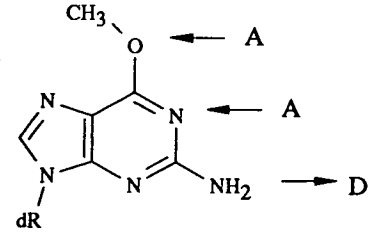
N(7)-Methyldeoxyguanosine.



Deoxyguanosine.



O^6methyl-2'-deoxyguanosine (distal).



O^6methyl-2'-deoxyguanosine (proximal).

D = Hydrogen bond donor.

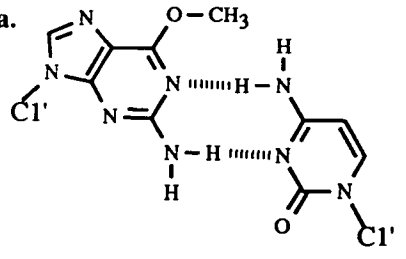
A = Hydrogen bond acceptor.

For the O6MedG derivative however, methylation effectively "freezes" the molecule into the "enol tautomer" such that the N(1) hydrogen bond donor site has been converted to an acceptor site. In addition, the orientation of the methyl group can be either distal, *syn* to N(1) which would block the Watson / Crick sites, or proximal, *anti* to N(1), which would not (Parthasarathy 1986 and Yamagata 1988). X-ray crystallographic data of the free base indicates that a distal orientation is preferred and indeed this proposition is supported by energy calculations which indicate that the molecule becomes thermodynamically more stable, by 3kcal/mol, by adopting this geometry (Yamagata 1988). Ultraviolet denaturation studies on self complementary dodecamers (Gaffney 1984) where O6MedG was placed opposite dA, dG, dC and T, indicate that the O6MedG.T pairing is the least stable of the pairs and O6MedG.dC the most stable. All are however considerably less stable than Watson / Crick pairs. This agrees to some extent with 2D-NMR studies of dodecamers containing O6MedG opposite dC and T respectively (Patel 1986a and 1986b). Figures 1.25a and 1.25c show the postulated pairings, with the methyl group adopting a distal orientation in both cases. ^{31}P NMR of the above duplexes indicate disruption of the phosphodiester backbone in the O6MedG.dC 12-mer but not in the O6MedG.T 12-mer. This is very interesting since it agrees with theoretical results which indicate that an O6MedG.dT pairing is possible at the normal C1'-C1' distance of 1.07nm whereas an O6MedG.dC base pair is only possible at the unusually large separation of 1.17nm between the C1'-C1' (Nagata 1988) and may indicate the presence of an O6MedG.dC wobble pairing.

The base pairs shown in Figures 1.25a to 1.25d, have been proposed from the 2-D NMR and mononucleotide X-ray crystallographic experiments discussed above. Both pairings with dC (1.25a and 1.25b) create a wobble base pair whereby the

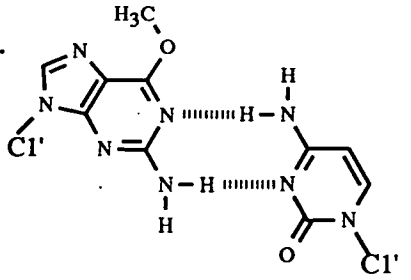
Figure 1.25.

1.25a.



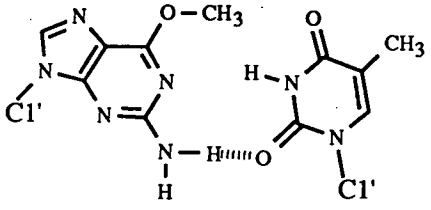
O^6 MedG (distal) : dC

1.25b.



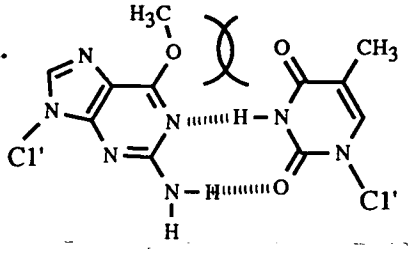
O^6 MedG (proximal) : dC

1.25c.



O^6 MedG (distal) : T

1.25d.

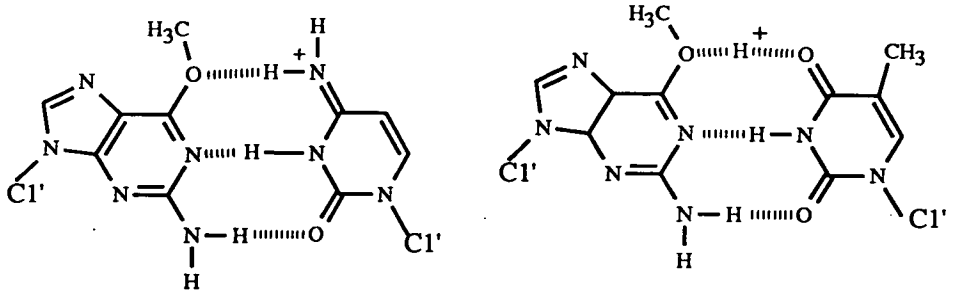


O^6 MedG (proximal) : T

dC is rotated slightly and moved into the minor groove in order to accommodate the hydrogen bonding. It seems unlikely that the orientation of the methyl group will affect the pairing. With T, (Figures 1.25c and 1.25d), the situation is slightly more confused as it is not clear how the carbonyl oxygen lone pairs of thymine and guanine will interact. Figure 1.25c has only one hydrogen bond which is a very unlikely situation since two hydrogen bonds are normally required for the overall "cooperativity" of the base pair (Saenger 1984) and therefore of the duplex. If however there is no great repulsion between the oxygen lone pairs, (Figure 1.25d), then a non-wobble base pair can form. This would be consistent with the evidence from the ^{31}P NMR (Patel 1986b). Watson / Crick hydrogen bonding can be proposed for the base pairs in Figure 1.26, whereby the bases utilise what would be their normal Watson / Crick sites.

The only efficient form of repair of this lesion is demethylation of the O^6MedG by the enzyme O^6-methyl guanine methyl transferase to regenerate guanine (Demple 1982 and Graves 1989). This enzyme is similar in both mammals and bacteria,

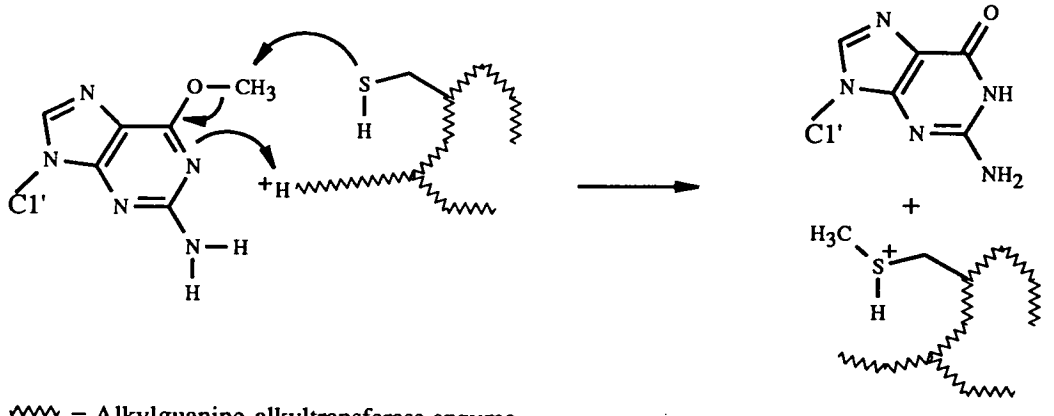
Figure 1.26.



Protonated base pairs. O^6MedG.dCH$^+$. and O^6MedG.TH$^+$.

indicating its early evolution. O^4-Methylthymine is also a reactive substrate for the enzyme which repairs the lesion by transferring the alkyl group onto one of its cysteine residues thereby repairing the mutagenic lesion and restoring the fidelity of replication, (Figure 1.27). Thus O^6-methyl-guanine methyl transferase is an inducable suicide enzyme (Sciccitano 1986).

Figure 1.27.



~~~ = Alkylguanine-alkyltransferase enzyme.

1.5.2. Formation of 7-hydro-8-oxo-2'-deoxyguanosine.

Radicals are utilised extensively in biology for catalysis by enzymes such as isopenicillin-N-synthetase and ribonucleotide reductase (Stubbe 1988) and by phagocytes (white blood cells) which produce peroxide, hypochlorous acid and superoxide to destroy bacteria that they recognise as foreign (Aruoma 1991). The
~~~

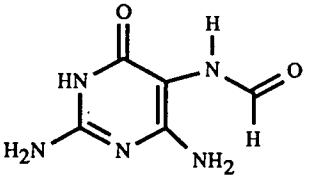
use of such systems has necessitated very strict control because of the destructive nature of these chemicals towards proteins, DNA and tissue etc. Consequently a variety of enzymes have evolved which metabolise radical oxygen species, for example superoxide dismutase, catalase and glutathione (McBride 1991) and they limit the damage caused by the daily dose of such agents. There are also a number of smaller molecule antioxidants such as ascorbic and uric acid, which are very reactive towards oxygen radical species (Rose 1990) but which can also themselves create adverse effects *in vivo* (Kasai 1984a). Overdoses of superoxide or hydroxyl radicals can occur through radiation (Dizdoraglu 1985) or exposure to chemicals (Kasai 1984b) which can swamp the natural defences. In addition to DNA damage, extensive tissue degradation can lead to the onset of diseases such as atherosclerosis and rheumatoid arthritis (Aruoma 1991). Hydroxyl radicals are extremely damaging to DNA and their ability to diffuse into the duplex can result in extensive strand scission and base modification. This reactivity has been utilised by a technique known as DNA footprinting whereby a pendant Fe(II)-EDTA group is attached to drugs or oligonucleotides such that site specific DNA splicing can be induced by the addition of a reducing agent such as dithiothreitol to the sample (Moser 1987). Therefore presence of trace amounts of transition metals, iron and copper in particular, *in vivo* can facilitate the production of hydroxyl radicals through Fenton and Haber-Weiss reactions.

Essentially. $Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + HO^- + HO\cdot$

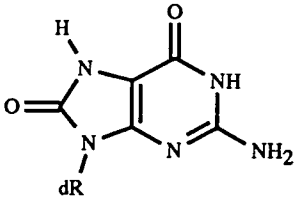
Such a system is employed in the Udenfriend system (Kasai 1984a) in which 8OxodG is produced in small amounts by the formation of radicals from a mixture of Fe^{2+}, ascorbic acid and hydrogen peroxide. Other known inducers of this deoxyguanosine derivative are aminophenols and asbestos (Kasai 1984b and 1984c). Although radical scavengers such as mannitol and thiourea protect against

damage to some extent, no real mechanism for hydroxyl radical metabolism has been identified. The two major guanine adducts formed by hydroxyl radical attack are 2,6-diamino-4-hydroxy-5-formamidopyrimidine and 8OxodG, (Figure 1.28.), and both are thought to play a role in the formation of tumours (Mallins 1990). Damage of this nature is being constantly repaired, and an endonuclease which specifically cuts out 8OxodG replacing it with dG has been discovered in a variety of mammals (Chung 1991). Thus the initial damage decreases with time after the attack. 8OxodG residues are then excreted in the urine, with an average daily excretion of around 178 residues each day per cell. The electrochemical activity of 8OxodG can be utilised in the routine analysis of human urine samples as an indicator of oxidative damage (Shiginaga 1989) .

Figure 1.28.



2,6-Diamino-4-hydroxy-
5-formamidopyrimidine.



8-Oxo-7-hydro-2'-deoxyguanosine.

Structural studies have been carried out which have established that the predominant tautomer of 8OxodG is the 8,6-diketo derivative (Culp 1989, Aida 1987 and Kasai 1987), and not the 8-hydroxy- tautomer as had been previously thought. In addition, as is very common with C-8 substituted purines, the free nucleoside prefers to adopt a *syn* glycosidic orientation (Cho 1990 and Uesugi 1977), thus it has the potential to act as a thymine analogue and pair with deoxyadenosine or to remain *anti* and pair with deoxycytidine. Replication experiments using 8OxodG in place of dG (Shibutani 1991) have established that dA and dC are incorporated preferentially, but that the ratio varies with the

polymerase used. Most striking however is the observation that chain extension beyond the lesion is preferential for the dA.8OxodG pair and only occurs at a much reduced rate beyond dC.8OxodG. This is in contrast to earlier work, (Kuchino 1987) which indicated that gross misreading occurred at the 8OxodG site and the flanking bases. The insertion of deoxyadenosine and deoxycytidine opposite 8OxodG is consistent with the results of the NMR analysis of the following two sequences:-

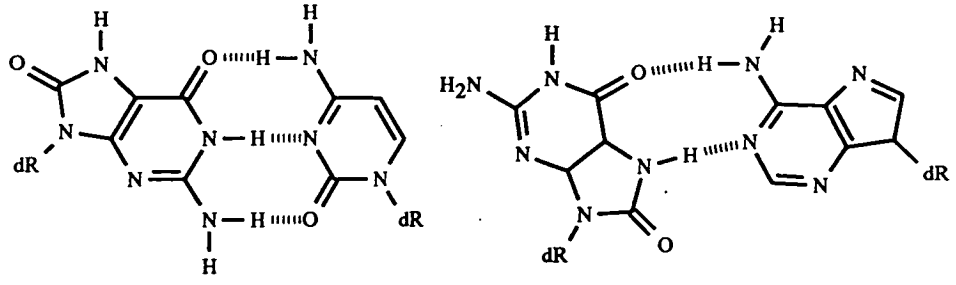
Sequence A) d[CGC**8OxoG**AATTAGCG]

Sequence B) d[CGC**8OxoG**AATTCGCG]

Both sequences are self complementary with sequence A yielding a 8OxodG*(syn)*.dA*(anti)* pair (Kouchakdjian 1991) and Sequence B a 8OxodG*(anti)*.dC pairing (Oda 1991). These results indicate the presence of stable base pairs which are accommodated symmetrically into the duplex with no apparent disruption of the phosphodiester backbone. The pairings established, (Figure 1.29), are in complete agreement with the replication experiments of Shibitani, and give an accurate explanation as to the reasons for the incorporation of dA and dC. The 8OxodG*(anti)*.dC pairing observed in sequence B, (Figure 1.29a), is consistent with a Watson / Crick dC.dG pairing and in this instance is not considered a mutagenic lesion, indeed the melting temperature of the modified sequence is 44°C, only 3 degrees lower than the native under the same conditions (Oda 1991).

In contrast, the 8OxodG*(syn)*.dA*(anti)* pairing observed in sequence A, (Figure 1.29b), is similar to the dG*(syn)*.dA mismatch discussed earlier although it would be expected to be a much more stable duplex since protonation of the deoxyadenosine is not required, thus allowing the formation of a base pair with

Figure 1.29a. **1.29b.**



8oxodG . dC 8oxodG . dA

two hydrogen bonds at physiological pH, resulting in a dG.dC to T.dA transversion mutation. Although in the free nucleoside the *syn* orientation is preferred, when helical constraints are placed on the mismatch, the conformation adopted will be dependent upon the nature of the opposing base. It is not so easy to rationalise the result that chain extension is only observed when a 8OxodG.dA pair is formed. However from *ab-initio* calculations it has been established that introduction of an oxygen at the C-8 position influences the entire electrostatic potential of the molecule, giving it a negative character (Aida 1987). This will undoubtedly have a major effect on the base stacking interactions and in addition could drastically alter the DNA / DNA polymerase interaction. It may be that 8OxodG*(anti)*.dC forms a very unusual enzyme / substrate interaction, which does not occur with 8OxodG*(syn)*, such that the base may not be in a favourable base stacking environment. This may be recognised by the proofreading domain as a mismatch preventing further synthesis. Variations in polymerase specificity between species may account for the differences observed in base incorporation opposite the lesion when using different DNA polymerase enzymes.

Aims.

In order to understand fully the potential of modified bases to induce mispairing and hence the formation of altered DNA sequences, it is important to understand at a molecular level the way in which such bases form mutagenic lesions in duplex DNA. Factors such as base pairing and stacking, symmetry of the resultant pair and introduction of new functionalities into the minor and major groove are all important in determining whether modifications are tolerated or not. Study in this field is extremely important, primarily because DNA mutation, as a result of nucleotide modification, is known in many cases to result in the onset of cancers.

Two modifications of dG, O^6MedG and 8oxodG, are known to induce base mispairs which can result in insertion of an incorrect amino acid during translation of the gene. The exact mechanism for the pairing of O^6MedG and 8oxodG is not fully understood and the aim of this thesis was to prepare oligonucleotides containing one of the above modified bases and then to characterise the base pair mutation by X-ray crystallography and ultraviolet melting.

Work on O^6MedG focused entirely on its effect on duplex DNA since a suitable precursor could be purchased, thus allowing its trouble free incorporation into oligonucleotides. A suitable route for the synthesis of 8oxodG was not however well established and originally the aim was to synthesise a suitable precursor from deoxyguanosine. As a result of some low yielding reactions the study was altered to encompass an investigation of some C(8) modified dG nucleosides and their stability in DNA. Thus novel functionalities, eg. bromine and p-nitrophenylthio, can be introduced into DNA at a site which protrudes into either the major or minor groove. Such modifications are potentially useful for DNA / protein crosslinking and studying the tolerance of different functionalities with regard to protein / DNA recognition.

Chapter 2.

2.1 Introduction.

The initial stages of chemical carcinogenesis frequently involve the interaction of genotoxic agents with DNA to produce covalent modifications in the form of DNA adducts (Miller 1981, Searle 1986 and Ames 1979). An important example of this is the alkylation of the O^6 position of guanine residues in DNA resulting from exposure to methylating agents such as N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) (Karran 1982), methyl methanesulphonate (MMS) and N-methyl-N-nitrosourea (MNU) (Saffhill 1985). The presence of O^6MedG constitutes a mutagenic lesion which is known to specifically induce G.C to A.T transition mutations (Loechler 1984) and it has been established that proto-oncogenes can be converted to oncogenes by such a process (Mitra 1989). Hence the formation of the O^6MedG.T base pair during replication can give rise to a carcinogenic lesion (Reddy 1982 and Zarbl 1985). In recent years the biochemical processes involved in chemically induced carcinogenesis have been studied in considerable depth. However, in order to understand further the mechanisms of mutagenesis, it is necessary to analyse precisely the molecular details of the lesions produced when genotoxic agents interact with DNA. With this overall objective in mind, we have determined the structure of such a lesion, the O^6MedG.T base pair in a B-DNA duplex.

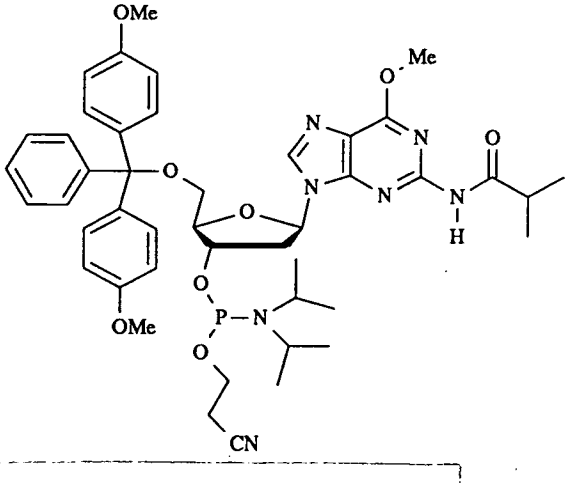
2.2. Results and discussion

1.1a. d[CGC<u>O6MeG</u>AATT<u>C</u>GCG].

1.1b d[CGC<u>O6MeG</u>AATT<u>T</u>GCG].

The above oligonucleotides were synthesised using the O6MedG phosphoramidite, (Figure 2.1). Synthesis and deprotection are described in the experimental section. After synthesis and purification, the above oligonucleotides were subjected to snake venom digest analysis to establish the nucleoside composition of the oligonucleotide and to confirm the presence of the O6MedG moiety.

Figure 2.1.



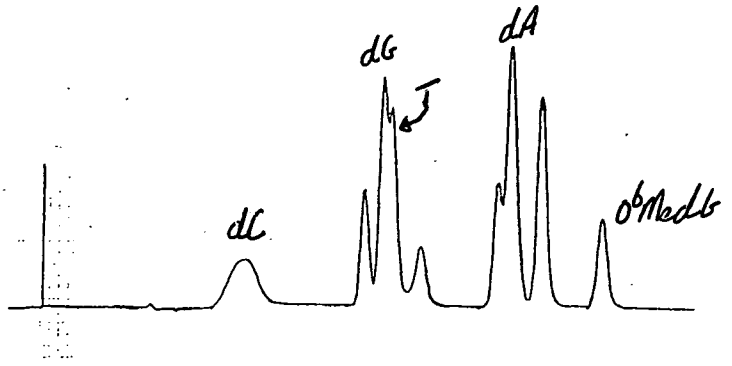
O6MedG phosphoramidite

2.2.1. Snake venom digest analysis.

Enzyme digest analysis allows the base composition of an oligonucleotide to be determined by degrading the polymeric chain into its constituent nucleosides. The sample can then be analysed by HPLC, using suitable conditions for the separation of the nucleosides.

Figure 2.2a.

Method 1.



2.2b.

Method 2.

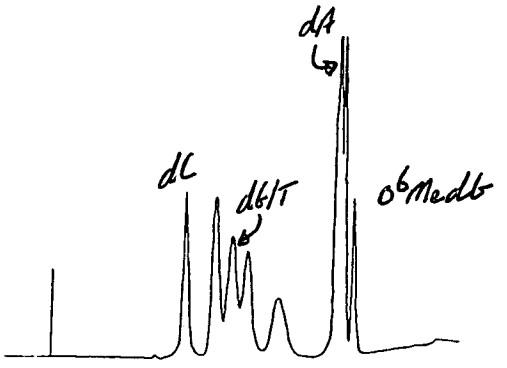


Figure 2.3a.

Method 2 digest of **1.1a.**

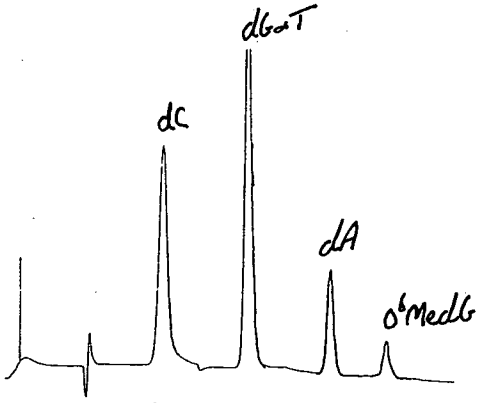
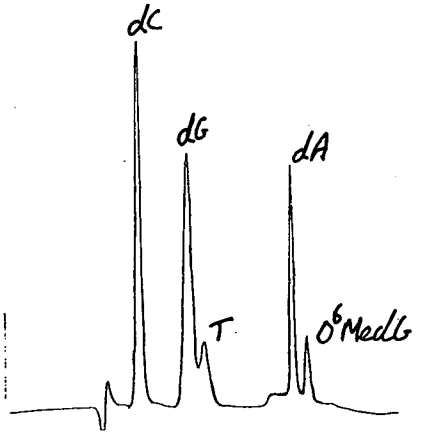


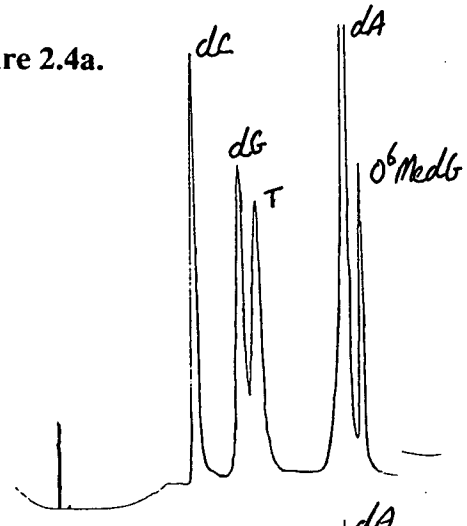
Figure 2.3b.

Method 1 digest of **1.1a.**

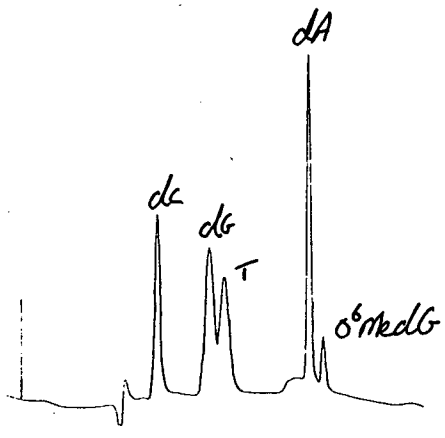


The separation of a selection of nucleosides, using two reverse phase HPLC techniques, can be seen, (Figure 2.2). Although neither system separates all the peaks, the HPLC conditions can be optimised depending on the nucleoside composition of the oligonucleotide. The main drawback of method 2 is that dG and T are not well resolved when dG is in greater concentration, e.g digest of **1.1a.** (Figure 2.3a), only elutes as four peaks. Method 1 in contrast separates dG and T (difference in retention time of 1 minute), (Figure 2.3b), thus allowing the base composition of the oligonucleotide to be more accurately assigned. Therefore, since the separation of dG and T is important for the accurate analysis of the oligonucleotides, method 1 was used. The HPLC traces shown, (Figure 2.4), indicate the elution profile of the standards, (Figure 2.4a) with the base composition of oligonucleotides **1.1a** and **1.1b** shown in Figures 2.4b and 2.4c.

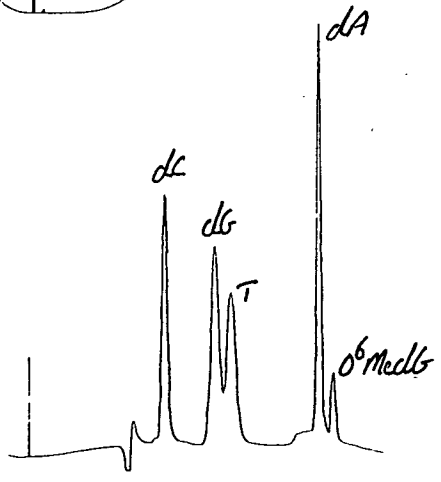
Figure 2.4a.



2.4b.



2.4c.



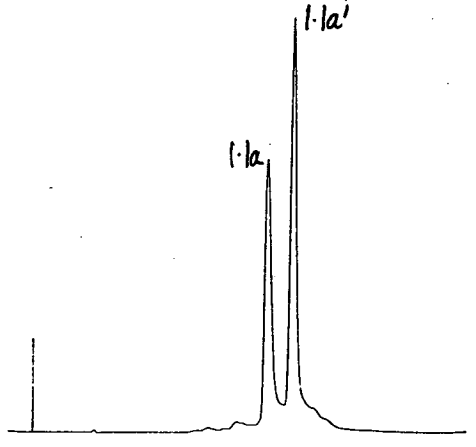
Although only qualitative results were obtained from this work it demonstrated that the oligonucleotides were pure, had the correct base composition and that the O^6MedG residue had survived the conditions of oligonucleotide synthesis and purification.

2.2.2 HPLC analysis.

Analysis of the deprotecting dodecamer **1.1a.** after 4 days deprotection, revealed two very distinct peaks on the elution profile (Figure 2.5a) with retention times of 16.9 min. (peak **1.1a**) and 18.4 min. (peak **1.1a'**).

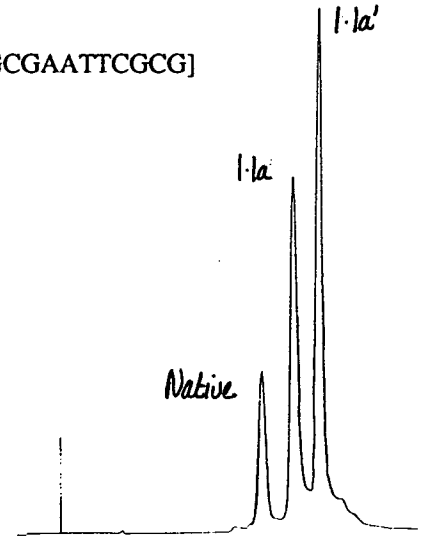
Figure 2.5a.

1.1a. Four days deprotection.



2.5b.

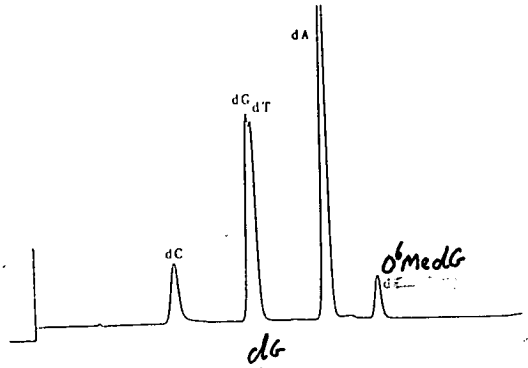
Mixed injection. **1.1a.** and d[CGCGAATTCGCG]



A mixed injection with the native sequence, d[CGCGAATTCGCG], (Figure 2.5b) showed elution of three peaks, with the native oligonucleotide eluting *c.a.* 2 min. before the modified oligonucleotides, thus confirming that no demethylation of the O6MedG moiety to dG had occurred. Peaks **1.1a** and **1.1a'** were then collected separately, reinjected as analytical samples and found to elute as single peaks, thus discounting hairpin loop formation or aggregation of the oligonucleotide **1.1a.** confirming that two different oligonucleotides were present.

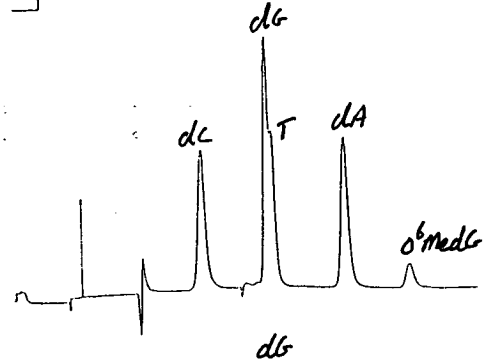
Figure 2.6a.

Base standards.



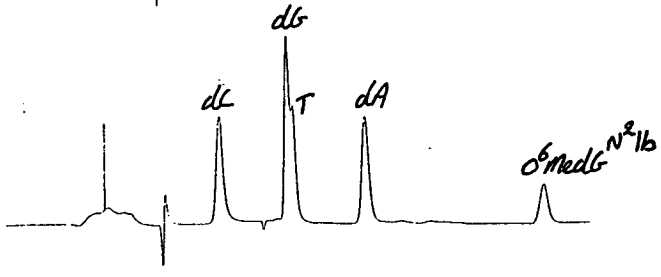
2.6b.

1.1a. enzyme digest.



2.6c.

1.1a' enzyme digest.

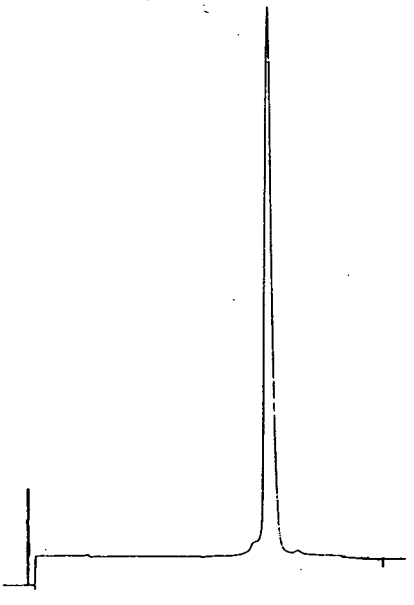


Both samples were then digested with snake venom and by comparing their elution profiles with known standards, (Figures 2.6a-c), the identity of the two samples was established. Although the peak assigned as $O^6MedG^{N(2)Ib}$ could not be positively identified, because no $O^6MedG^{N(2)Ib}$ was available, it was later confirmed as this by subjecting the two peaks to the oligonucleotide deprotecting conditions for a further week, Figures 2.7a and 2.7b. **1.1a** still eluted as one component with the same R_t, whereas the profile of **1.1a'** was the same as in Figure 2.5a, indicating that further deprotection had taken place.

The above analysis clearly demonstrated the difficulty in removing the N^2-isobutyryl moiety from O^6MedG and is consistent with previous reports, (Kuzmich 1983). Consequently, a 2 week deprotection strategy was utilised which resulted in a deprotected oligonucleotide mixture with a major and minor product ratio of 9:1, (Figure 2.8a). Subsequent purification yielded the pure dodecamer as a single peak by HPLC, (Figure 2.8b).

Figure 2.7a.

1.1a. after one week further deprotection.



2.7b.

1.1a' after one week further deprotection.

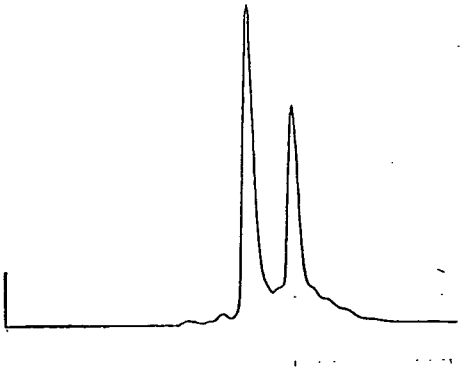
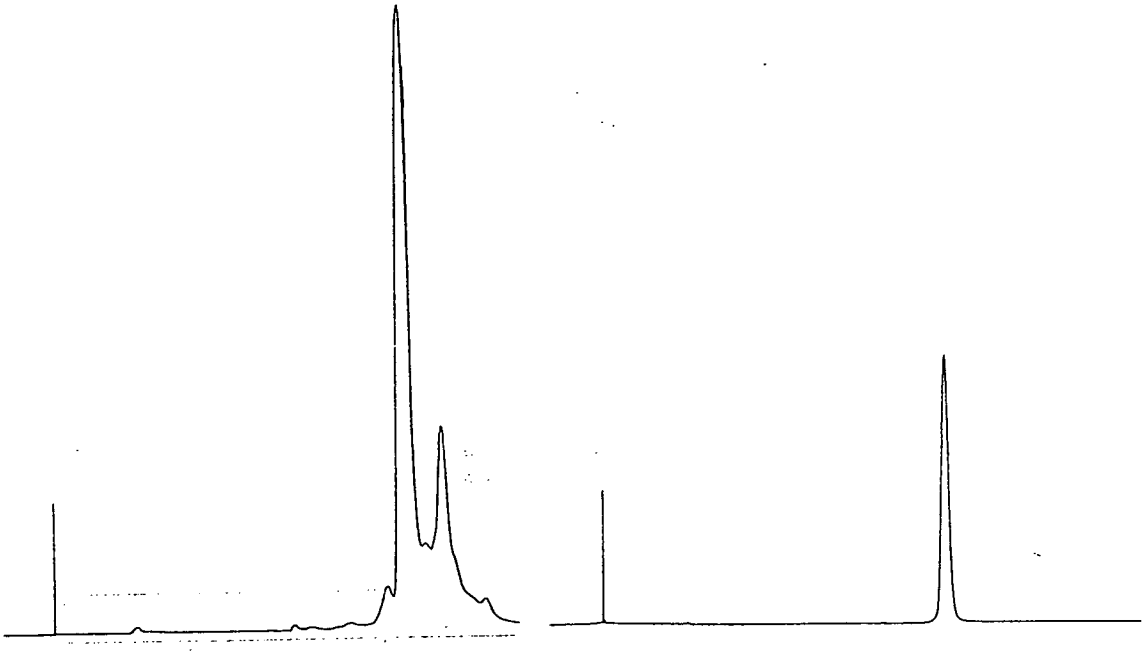


Figure 2.8a.

1.1b. after two weeks deprotection.

2.8b.

1.1b. after HPLC purification.



This careful analysis of all oligonucleotides ensured that the samples were not contaminated by failure sequences or oligonucleotides containing degraded bases. Thus oligonucleotides containing O^6MedG, (**1.1a.** and **1.1b.**), were obtained in very high purity, which is crucial both for the accurate determination of melting temperatures and the formation of high quality crystals necessary for X-ray diffraction studies.

2.2.3. Ultraviolet melting determination.

The technique of ultraviolet melting relies on the fact that double stranded DNA has a lower ultraviolet absorbance than single stranded DNA. Therefore, if the thermal denaturation of a self complementary dodecamer is carefully monitored, at 260 nm, a rise in the absorbance (ca. 25%) is observed as the duplex denatures to yield single strands. Duplexes held together with Watson / Crick bonding denature at higher temperatures, d[CGCGAATTCGCG] at *ca.* 45°C, than similar sequences containing base mispairs, d[CGCO6MeGAATTCGCG] *ca.*23°C.

2.2.3a. Concentration dependent UV melting.

According to Equation 1, determining the melting temperature, T_m, of a duplex under carefully controlled conditions allows thermodynamic parameters to be calculated which define the stability of double helical DNA. (Marky 1987)

Equation 1. $\mathbf{T_m^{-1} = R.\ln C_t / \Delta H^\phi + [\Delta S^\phi - \ln(1.204).R] / \Delta H^\phi}$

Therefore a plot of T_m^{-1} versus $\ln C_t$ will yield a straight line of gradient $R / \Delta H^\phi$ which bisects the y-axis at $[\Delta H^\phi - \ln(1.204).R] / \Delta H^\phi$, thus enabling the calculation of ΔH^ϕ, ΔS^ϕ and ΔG^ϕ. This analysis was carried out on dodecamers **1.1a.** and **1.1b.** with the results, in comparison to the native and other mismatched oligonucleotides, shown in Table 3 and Figure 2.9.

These results indicate that both dodecamers have almost the same duplex stability and therefore the preference of O6MedG to pair with T cannot be a result of favourable thermodynamic stability.

Figure 2.9.

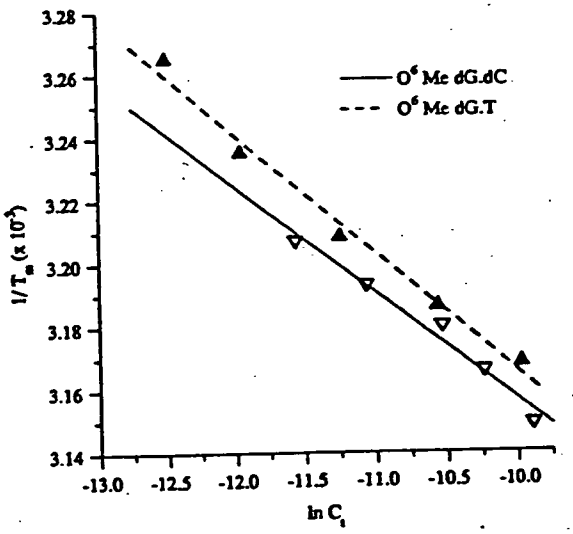


Table 2.3.

Thermodynamic Parameters for d[CGCXAATTYGCG]$_2$ duplexes at pH 7.0 in 1.0 M NaCl / 10 mM sodium phosphate / 1 mM EDTA.

Base pair.	ΔHϕ kJ.mol^{-1}	ΔSϕ J.mol^{-1}.K^{-1}	ΔGϕ kJ.mol^{-1}	T_m, K
dG.dC	-430.1	-1164	-83.2	344.6
dA.T	-398.8	-1084.7	-75.7	341.6
dG.T	-331.1	-937.6	-52.3	324.4
O^{6}MedG.dC	-213.6	-592.3	-37.2	317.1
O^{6}MedG.T	-208.4	-577.3	-36.4	315.2

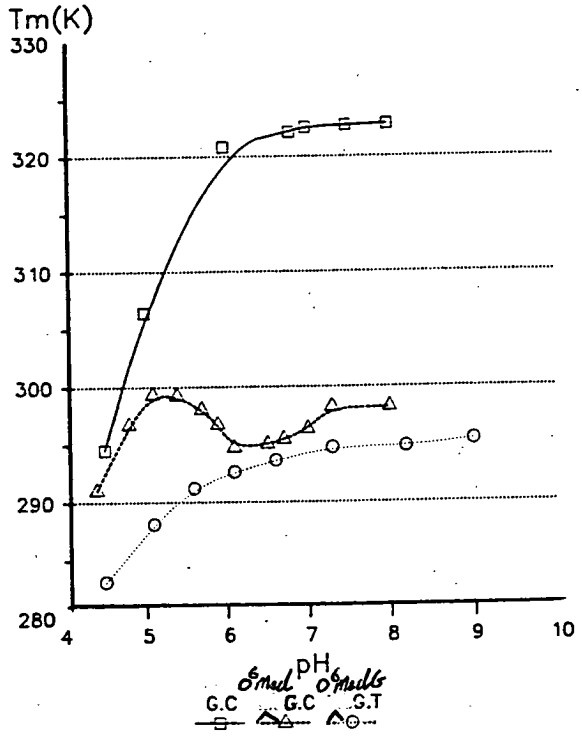
2.2.3b. pH dependent UV melting.

Another study was carried out using dodecamers **1.1a.** and **1.1b.** by monitoring T_m as a function of pH. Establishing the variation in oligonucleotide stability as a function of pH determines whether any protonated or deprotonated base pairs are involved in hydrogen bonding. The results from this study are shown in Figure 2.10.

Before carrying out these pH dependent studies the chemical stability of O^{6}MedG at

low pH was determined in the following way: the oligonucleotide d[CGCO6MeGAATTCGCG] was dissolved in 0.1 *M* phosphate buffer at pH 5.0 and after two weeks reversed-phase HPLC analysis showed that there was no significant degradation. Mixed injections with the native sequence d[CGCGAATTCGCG], which elutes much earlier on reversed-phase HPLC, further confirmed that the O^6MedG containing oligonucleotide had not undergone demethylation.

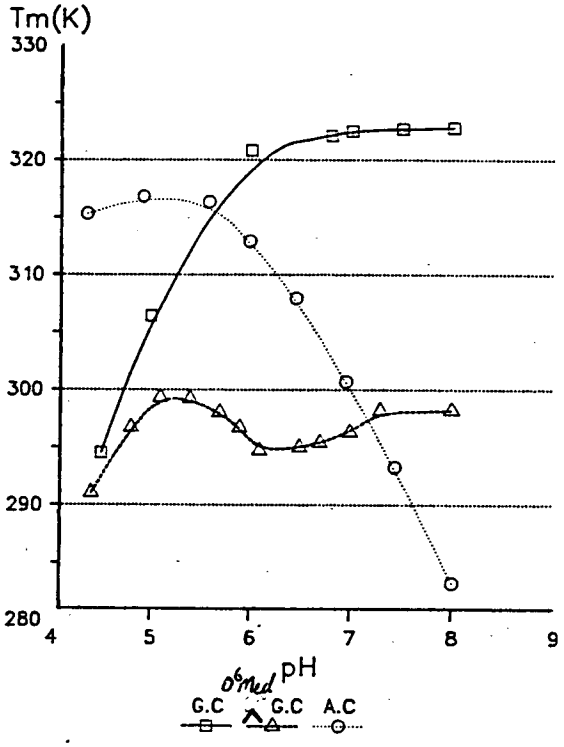
Figure 2.10.



In the above graph, the profile obtained for the two mismatch oligonucleotides, **1.1a** and **1.1b**, have been plotted alongside the profile of the control sequence, upper curve, which illustrates the pH dependent UV melting curve for a standard Watson / Crick duplex, which has no variation in stability between pH 5.8 and 8.0. Below pH 6.0 the melting temperature drops dramatically, a result of the protonation of dC-N(3) which causes destabilisation of the duplex. Although much more unstable than the native dodecamer, oligonucleotide **1.1b.** exhibits the same trend of pH dependency as the

native dodecamer. This is in stark contrast to the curve obtained from oligonucleotide **1.1a.** which demonstrates a variation in stability in the region between pH 5 and 8. The curve has a minimum point at pH 6.2 and appears to be a composite of two curves, (Figure 2.11.), one which displays a variation in stability between pH 6.5 and 8, as in the native dG.dC sequence, and one which displays a variation in stability between pH 5.3 and 6.5, as is found for the dA.dC mismatch.

Figure 2.11.

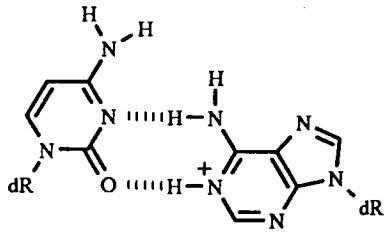


The dA.dC mismatch (Hunter 1987), Figure 2.12a, contained in the dodedecamer dCGCAAATTCGCG, forms a wobble pair with two hydrogen bonds, one of which requires protonation of dA-N(1). The duplex is very unstable at pH 7.7 because it is unable to exist in the non-protonated form, and as a result the duplex cannot maintain a double helical structure. At lower pH, N(1) protonation becomes possible such that two hydrogen bonds can form, and so the stability of the duplex increases to a maximum at pH 4.7. This may help explain the pH dependent behaviour of dodecamer **1.1a.** which

indicates the existence of a protonated and a non-protonated base pair. In basic conditions the curve parallels that of the dG.dC dodecamer implying a non-protonated base pair as in Figure 2.12b. At acidic pH however, a minimum point is reached (at pH 6.2) where the curve then begins to parallel that of the dA.dC dodecamer, reaching a maximum at pH 5.6. This is due to the protonation of dC-N(3) of the mispair, resulting in the pairing shown in Figure 2.12c. Since both pairings have equal stability it is reasonable to assume that in an *in vivo* environment an O^6MedG.dC pairing will exist in the non-protonated wobble form, Figure 2.12b.

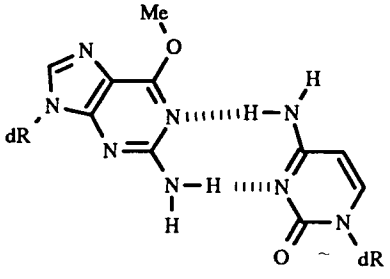
Figure 2.12.

12a.



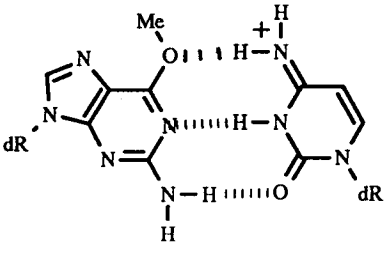
dC : dAH^+

12b.



Wobble pairing of O^6MedG.dC

12c.



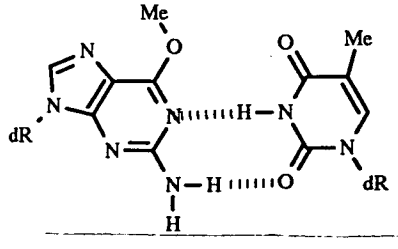
O^6MedG.dCH^+

2.2.4. Crystal structure of d[CGCO6MeGAATTTGCG]$_2$

The X-ray structure analysis of the self-complementary dodecanucleotide d[CGCO6MeGAATTTGCG]$_2$ shows that in overall shape the O6MedG.T mispair is remarkably similar to a Watson / Crick base pair. The bases are directly opposite each other, the glycosyl linkages are related by a pseudodyad, and in the minor groove the base pair is indistinguishable from a dG.dC base pair. Overall, there are very few structural differences between the O6MedG.T duplex and the well-studied native one, d[CGCGAATTCGCG]$_2$ (Wing 1980). The helical parameters, torsion angles and hydration pattern are almost identical and the presence of the mutagenic base pair does not distort the sugar-phosphate backbone. Despite these considerable similarities, ultraviolet melting studies indicate that the O6MedG.T duplex is much less stable than the native dG.dC duplex ($\Delta\Delta G\phi$ 46.8 KJ.mole^{-1}), (Table 2.3). Hence the O6MedG.T base pair has a powerful destabilising effect in B-DNA.

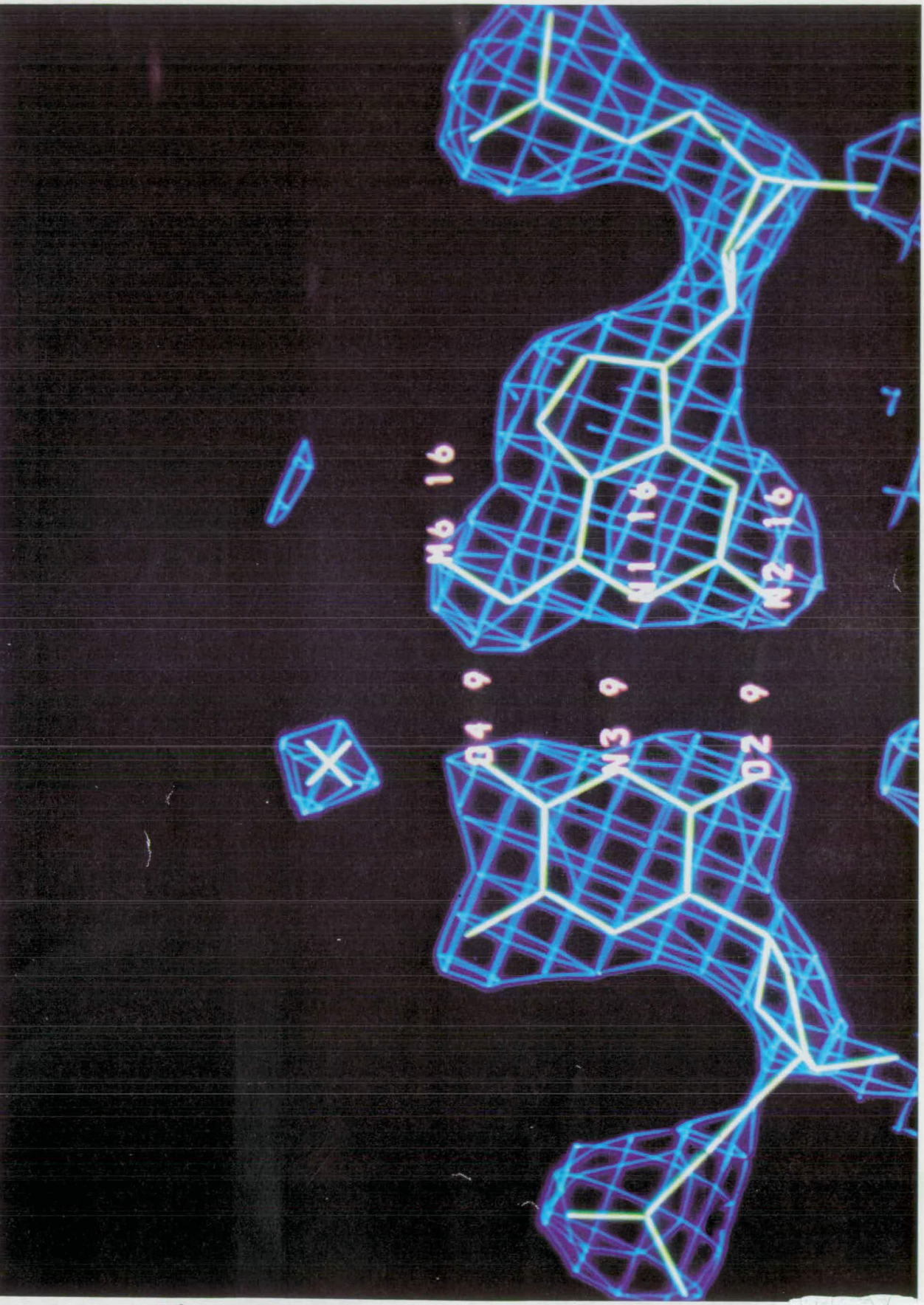
The two O6MedG.T base pairs in the dodecamer duplex are essentially identical and the $2F_o$-F_c map of the O6MedG(9).T(16) pair is shown in Figure 2.13, overleaf. There are three close contacts between the O6MedG and T bases; one of 2.9*A* from purine-O6 to pyrimidine-O4, a second of 2.9*A* from purine-N(1) to pyrimidine-N(3), and a third of 2.8*A* between purine-N2 and pyrimidine-O2. At this resolution hydrogen atoms were not observed and the most likely base pair is shown in Figure 2.14, with each base in its most common tautomer.

Figure 2.14.



O^6 MedG : T base pair

FIGURE 2.13

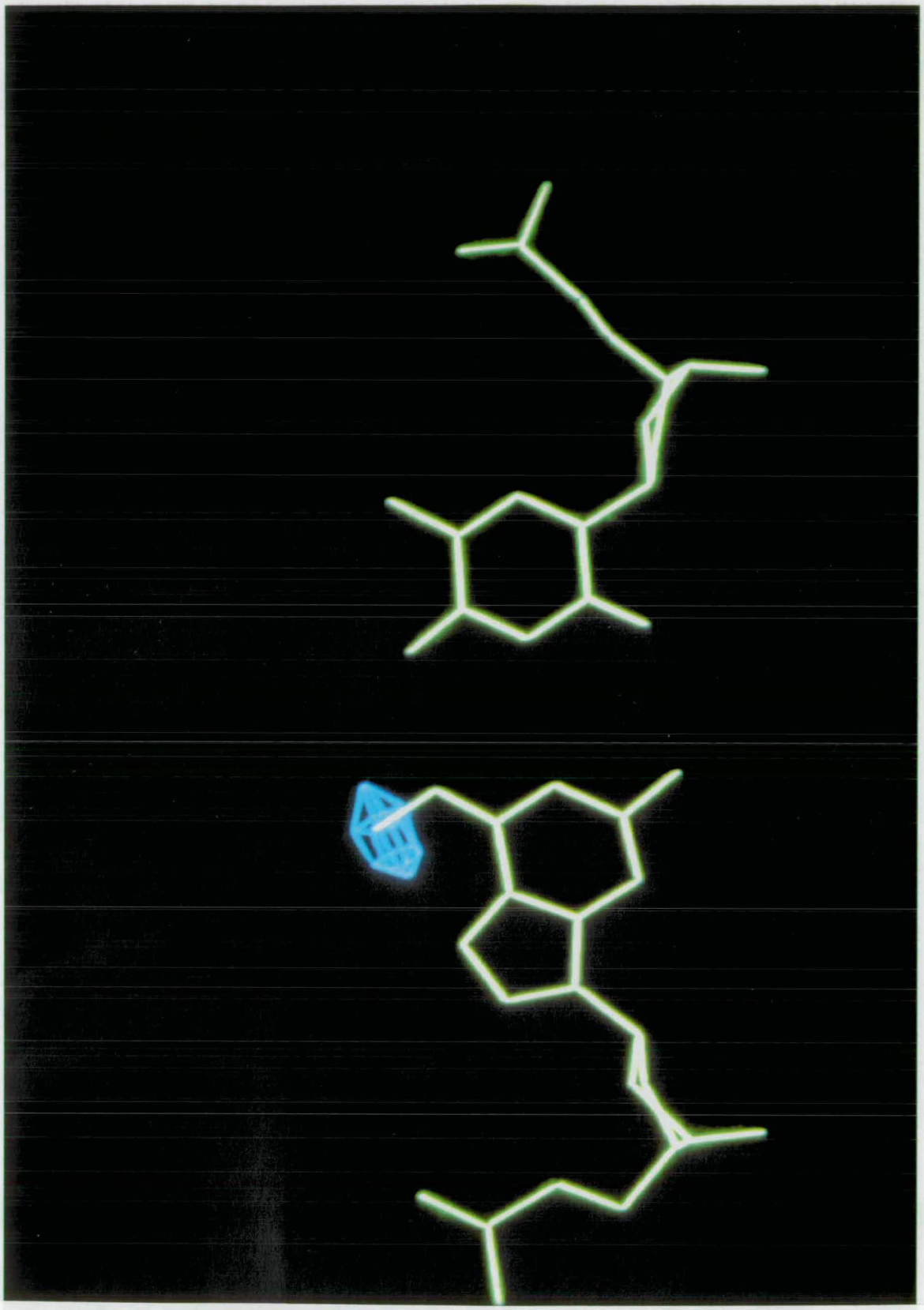


Electron density map showing the T:O^{6}MedG base pair.

There is unlikely to be a hydrogen bond between O^6-purine and O^4-pyrimidine, as neither has an attached hydrogen and neither functional group is sufficiently basic to be protonated at pH 6.3. In addition, there is no evidence in the pH-dependent ultraviolet melting profile of the duplex (Figure 2.10) for a protonated base pair. Hence, formation of the double strand from the fully hydrated single strands will lead to the loss of a hydrogen bond between each of these oxygen atoms and neighbouring solvent molecules. As these are not replaced by inter-base hydrogen bonds, the effect will be to destabilise the duplex. The interaction between purine-O^6 and pyrimidine-O^4 will be destabilising for the additional reason that the two electronegative oxygen atoms are forced together by the two adjacent strong hydrogen bonds in the base pair (Nagata 1988), one from pyrimidine-N(3) to purine-N(1) and a second from purine-N^2 to pyrimidine-O^2. It is possible to postulate other forms of the O^6MedG.T base pair by invoking minor tautomers but there is no direct experimental evidence for their existence. The base pair in Figure 2.14 has two hydrogen bonds which are adjacent to each other and the resultant co-operativity allows the formation of a stable base pair. The O^6MedG.T base pairing found in the present X-ray structure is identical to that postulated from molecular orbital and molecular mechanical calculations, and from an nmr study of a mixture of the O^6MedG and thymidine nucleosides in chloroform solution (Williams 1987). However it differs in important details from that proposed in an oligonucleotide nuclear magnetic resonance study, although interpretation of the spectra was limited by the lack of direct information on the relative orientation of the two bases, (Patel 1988).

The most striking differences between the O^6MedG.T mispair and a dG.dC base pair are in the major groove, owing to the presence of the methyl group attached to the purine O^6-atom, located proximal to the N(7)-atom in the plane of the purine ring (Figure 2.15). This will give rise to steric repulsion between the guanine base and the

FIGURE 2.15



Electron density map indicating the proximal orientation of the methyl group in the T:O^{6}MedG base pair.

attached methyl group. Although the distal conformation is preferred in the free nucleoside (Parthasarathay 1986 and Yamagata 1988), it is likely to be very unstable in the O6MedG.T base pair (Pedersen 1988) as it will prevent the formation of inter-base hydrogen bonds. In the proximal orientation the methyl group presents a steric barrier to any regulatory or repair enzyme that might otherwise interact with the guanine O6- or N(7)-atom. The appearance of the O6MedG.T base pair in the major groove is different from that of a Watson / Crick dA.T or dG.dC base pair in two additional details: **i)** there are two methyl groups, one attached to each base, and **ii)** there is no hetero-atom with a capacity to donate hydrogen bonds.

2.2.5. Conclusion.

When a guanine base in genomic DNA is converted to O6MedG by a chemical mutagen, the modified base normally codes for thymine instead of cytosine (Snow 1984), resulting in an *in vivo* mutation frequency of between 15% (Chambers 1985) and 75% (Bhanot 1986). The *in vitro* misinsertion frequency is more than 95% (Toorchen 1983) and the only efficient form of repair involves demethylation of the O6-atom of guanine by the enzyme O6-methyl guanine methyl transferase to regenerate guanine (Olsson 1980 and Demple 1982). Thus, the O6MedG.T base pair is recognised as being more similar to a Watson / Crick base pair than is the O6MedG.dC base pair. The reasons for this are unlikely to be thermodynamic in origin (Gaffney 1984 and 1989), as ultraviolet melting studies show that the duplex containing the O6MedG.T base pair is slightly less stable than the corresponding O6MedG.dC duplex at neutral pH ($\Delta\Delta G^{\circ}$ 0.8 KJ.mole^{-1}) (Table 2.3). Moreover, the O6MedG.T duplex is less stable than the duplex containing the dG.T wobble base pair mismatch which is rarely incorporated during replication owing to efficient proofreading ($\Delta\Delta G^{\circ}$ 15.9 KJ.mole^{-1}).

The incorporation of the O6MedG.T base pair in preference to the O6MedG.dC base

pair in genomic DNA can be rationalised in structural terms. The similarity in shape between the O6MedG.T mispair and a Watson / Crick base pair, particularly in the minor groove is striking, whereas the O6MedG.dC base pair has been postulated on the basis of theoretical and nuclear magnetic resonance studies to be a reverse wobble base pair as in Figure 2.12b (Patel 1986 and Pedersen 1988). This would, by analogy with mismatch base pairs, be removed by proofreading. Thus the enzymes in the cell nucleus responsible for DNA synthesis and repair discriminate in favour of the mutagenic lesion. We have studied the stability profile of the DNA duplex containing the O6MedG.dC base pair over a wide pH range (Figure 2.10) and have observed that the melting temperature falls slightly from pH 8.5 to pH 6.5, as would be expected for the unprotonated wobble base pair 2.12b. At lower pH however, there is a clear indication of duplex stabilisation, probably due to protonation of cytosine-N^3 of the mispair, which would lead to the formation of a Watson / Crick-like base pair, a resonance form of which is shown in Figure 2.12c. Thus, the O6MedG.dC base pair displays conformational flexibility. The protonated base pair 2.12c, which is probably present to some extent at neutral pH, would be expected on structural grounds to be incorporated during replication. Such a base pair has been identified in a nuclear magnetic resonance study of a mixture of the nucleosides in non-aqueous solvents (Williams 1987) and in the X-ray structure of a Z-DNA duplex (Ginell 1990). The occasional occurrence of 2.12c might explain why the presence of O6MedG in genomic DNA does not always lead to a mutation.

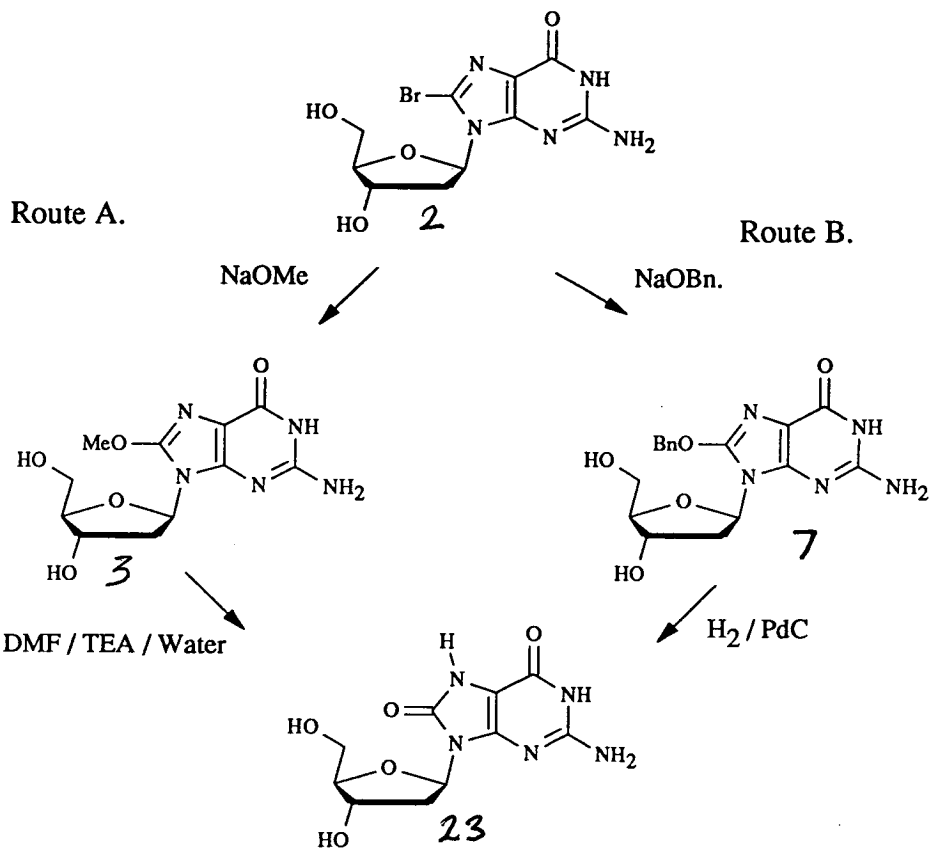
Chapter 3.

3.1. Introduction.

8-Oxo-7-hydro-2'-deoxyguanosine, (8OxodG), is a product of oxidative damage to DNA and can cause base mispairing during replicative events. The potential for 8OxodG to pair with either dC or dA is of great interest, since, as outlined in the introduction, (Figure 1.29.), pairing with dC has been shown in certain polymerase systems to halt replication, (Shibutani 1991), whereas pairing with dA would constitute a mutagenic lesion. In an effort to establish the base pairings possible with 8OxodG we have attempted its synthesis by a variety of methods utilising C(8)-modified dG derivatives both in the nucleoside form and on its incorporation into oligonucleotides.

8-Bromo-2'-deoxyguanosine, (8BrdG), is a very important intermediate in the synthesis of C(8)-dG derivatives since the presence of a bromine atom activates the C(8) position to nucleophilic attack and consequently enables the introduction of an alternative functionality, eg. methoxy or benzyloxy. It has been established from electron density calculations that the C(8) position of dG is the most susceptible to electrophilic and free radical attack, (Holmes 1964), thus facilitating the bromination. 8BrdG can be utilised as the starting material for two possible routes towards the synthesis of 8OxodG, (Figure 3.1). Route A involves displacement of the bromine atom by methoxide anion, (Tai-Shun Lin 1985), to produce 8-methoxy-2'-deoxyguanosine, (8MeOdG). The last reaction in this route was thought to be possible as a result of the work by Kuchino, (Kuchino 1987), who demonstrated that oligonucleotides containing 8MeOdG could be converted to their 8Oxo derivatives by the action of DMF / TEA / water.

Figure 3.1.



The alternative method, Route B, involves displacement of the bromine with benzyloxide anion. The resultant 8-benxyloxy-2'-deoxyguanosine[7](8BnOdG) can be easily transformed into the 8Oxo derivative *via* hydrogenation. (TAI-SHUN, 1985)

Thus, two alternative approaches exist, one which leads to the insertion of a protected 8OxodG (Route A, Sect. 3.2.) and the other an unprotected 8OxodG (Route B, Sect. 3.3.) into the oligonucleotide.

3.2. Conversion of 8MeOdG to 8OxodG.

(8MeOdG)

The crude product, **3**, was obtained as a pale yellow powder in approximately 80% yield and was used for subsequent reactions without further purification since all attempts at purification by crystallisation were unsuccessful despite reports to the contrary, (Tai-Shun Lin 1985). All attempts to repeat this work, and in alternative solvents, resulted in the formation of a very thick gel on cooling but no crystals. Analysis, (t.l.c. and HPLC), indicated no major impurities in the crude powder and it was therefore used in this form.

At no point was it possible to convert 8MeOdG to 8OxodG using the DMF / TEA / Water system reported by Kuchino, (Kuchino 1987). This reaction was also carried out on oligonucleotides containing an 8MeOdG base, where the pKa value of the N(1) proton is slightly higher, (Saenger 1984), but no reaction was observed, see section **3.2.3.**

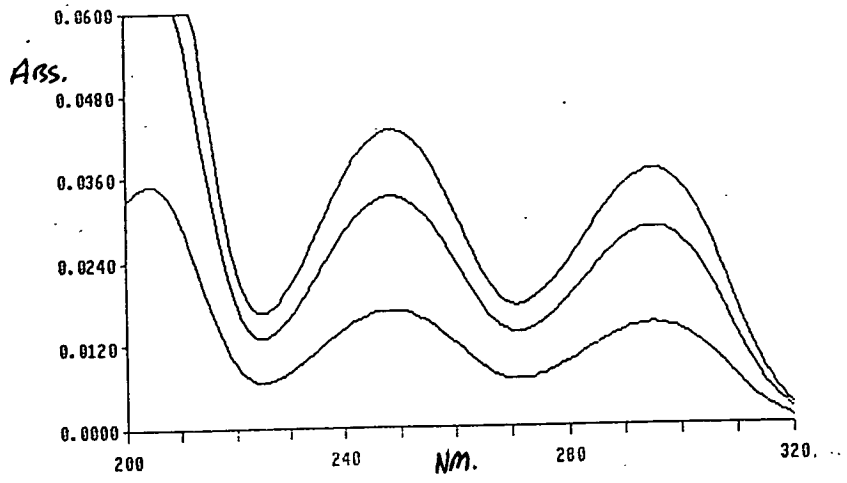
During an attempt to recrystallise 8MeOdG from a solution of pyridine / water (1:1) it was noticed that the sample had degraded into two distinct components (HPLC and t.l.c.) in equal quantities. Separation of the two components by HPLC, method 2, followed by UV analysis indicated that one component was 8MeOdG, (Tai-Shun Lin 1985) and the other 8OxodG, (Figure 3.2a), (Culp 1989). The presence of 8OxodG was further confirmed by co-elution, against standard 8OxodG, (HPLC, method 3.), (Figure 3.2b). Compound 23, See experimental p113.

Although the initial conversion in pyridine / water was not particularly high yielding, it was found that addition of TEA and DBU to the reaction mixture enhanced the conversion to nearly 100%, as determined analytically by HPLC and t.l.c. The use of thiophenol was also found to be beneficial to this transformation with the addition of TEA and DBU again enhancing the conversion to nearly 100%, (HPLC profiles

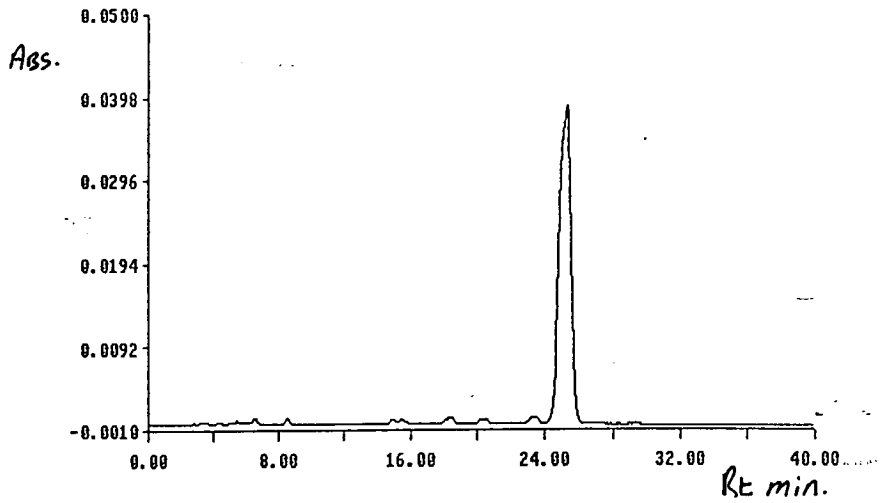
Figure 3.3a to 3.3d).

Figure 3.2.

a.



b.



Although it was clear from this work on the free nucleoside, that 8OxodG could be readily produced from 8MeOdG, a major problem was again the purification of the C(8)- alkoxy and oxo products. As a result it was decided to proceed with the incorporation of 8MeOdG into oligonucleotides and to carry out conversion after oligonucleotide synthesis was complete. This also had the advantage of allowing us to utilise the capacity of the synthesiser to deliver a solution of thiophenol / dioxan / TEA (20:78:2), to the column containing the oligonucleotide. This also simplifies purification, since excess thiophenol can be easily washed of the column using dry

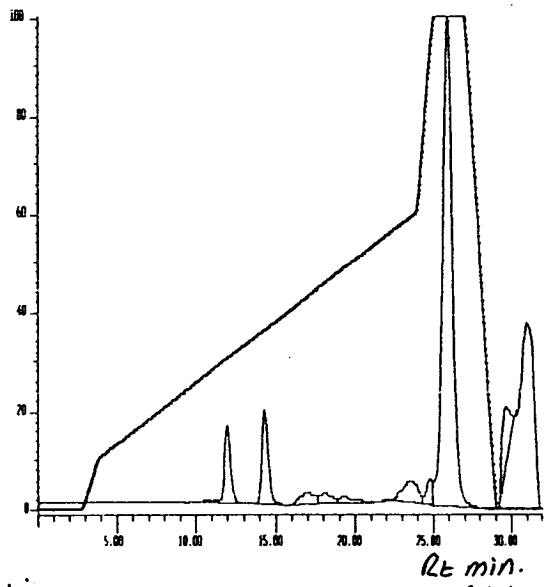
Figure 3.3.

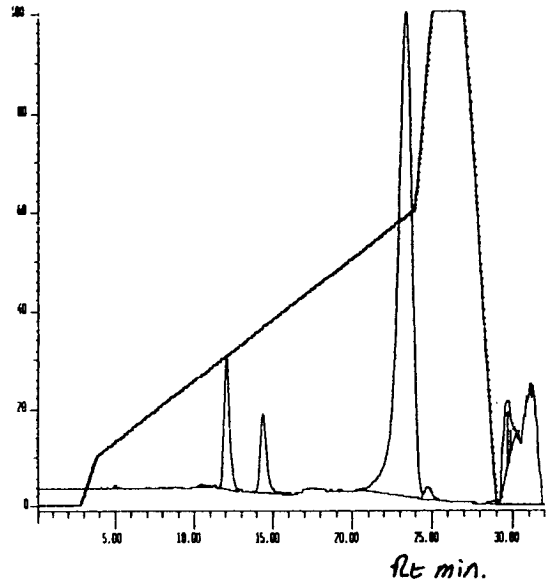
a).Pyridine/Methanol/Thiophenol.(Solution A)

Rt 8MeOdb = 14.5min.

Rt 80xodb = 12 min.

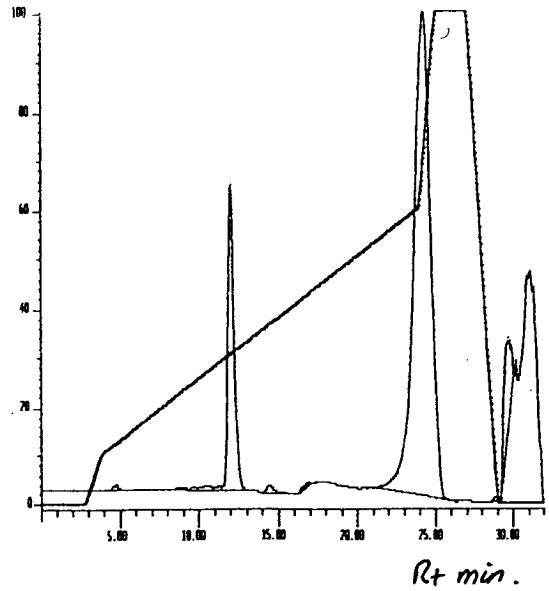
Solution
b). A + water.

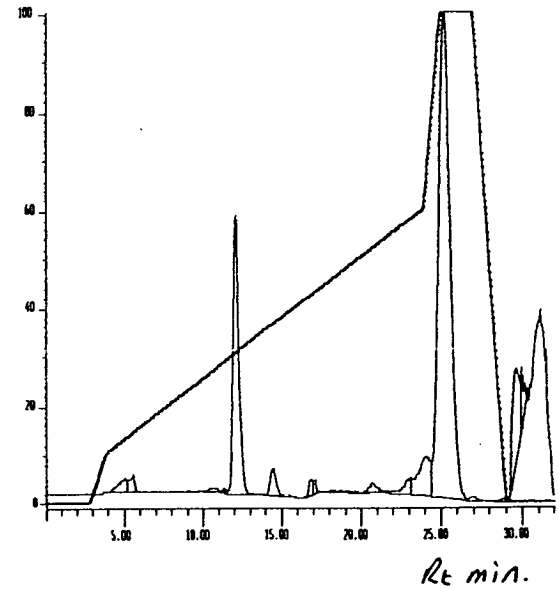




Solution
c). A + TEA.

Solution
d). A + DBU.



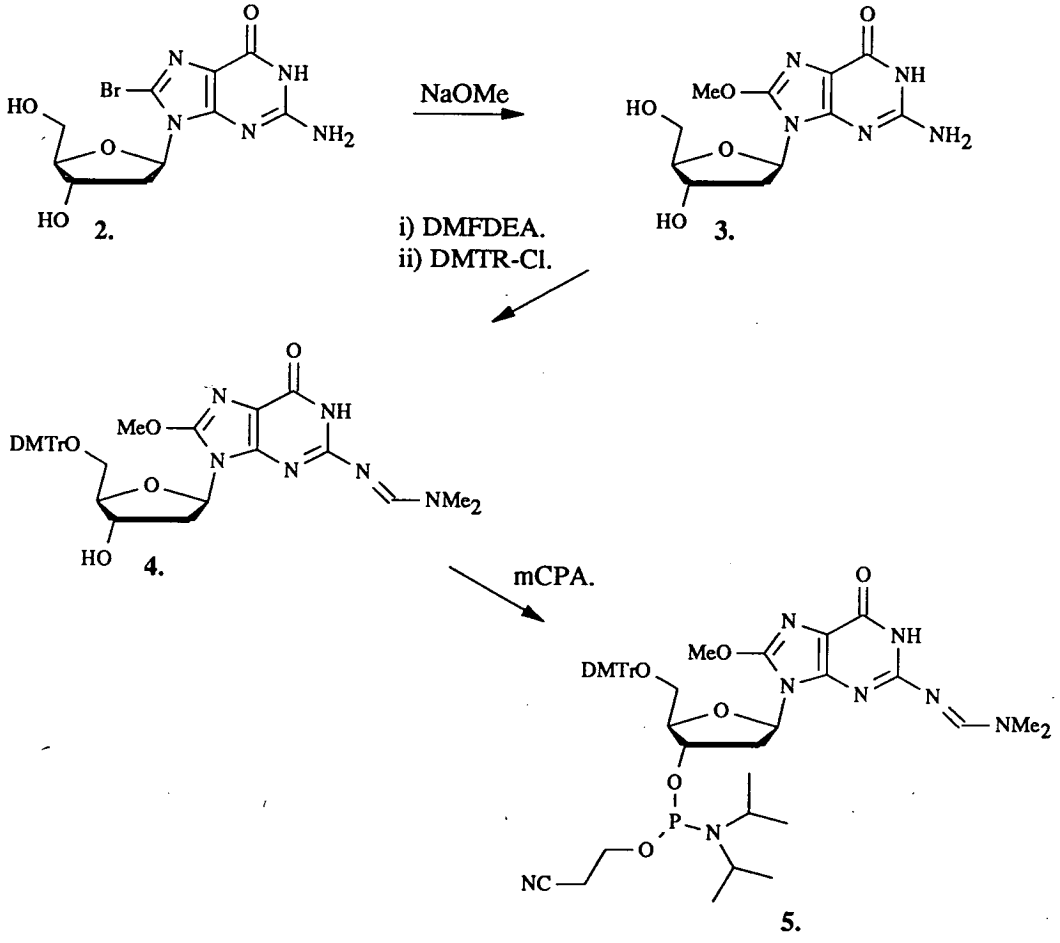


methanol. With this aim in mind, the synthesis of a suitable phosphoramidite monomer was achieved, and 8MeOdG subsequently incorporated into an oligonucleotide.

3.2.1. Synthesis of an 8MeOdG monomer.

Incorporation of 8MeOdG into oligonucleotides necessitated the synthesis of nucleoside monomer **5.** (Figure 3.4), which could be inserted into an oligonucleotide *via* the phosphoramidite strategy of DNA synthesis (Beaucage 1992).

Figure 3.4.



Having obtained compound **3.** in suitable quantity the requisite N^2 3'O- and 5'O-protecting groups were then attached.

The N^2 amino group of 8MeOdG was selectively protected by reaction of **3.** with dimethylformamide diethyl acetal to afford the N^2-methylene-dimethylamine derivative. With no purification, this was reacted with 4,4'-dimethoxytrityl chloride, which, in the presence of DMAP, reacts specifically on the primary 5'-hydroxyl moiety yielding compound **4.**, which exits as a mixture of *cis* and *trans* isomers. These isomers were easily visualised by t.l.c. and were shown to have identical chemical composition by FAB-MS. In addition, reaction of **4.** with 0.88 NH_3 removes the N^2 protection leading to the formation of only one trityl positive component at lower R_f. Reaction of compound **4.** with mCPA afforded the suitable phosphoramidite nucleotide derivative, **5.**, which was incorporated into oligonucleotides **6.1** and **6.2.** with a coupling efficiency of > 99%. The synthesis of this oligonucleotide therefore allowed a direct comparison to be drawn between formation of 8OxodG from 8MeOdG nucleosides, section **3.2.** and from 8MeOdG incorporated into DNA.

3.2.2 Oligonucleotides containing 8MeOdG.

Oligonucleotides **6.1.** and **6.2.**, see below, were synthesised as several 0.2 μM preparations and were subjected to the various reaction conditions found to promote the conversion of 8MeOdG to 8OxodG, section **3.2.**

6.1. d[CCCCC**8MeOGT**CCCCC]
6.2. Same sequence as **6.1.** but was retained on the CPG beads in its protected form. Reactions were carried out by injecting the reagents into the column.

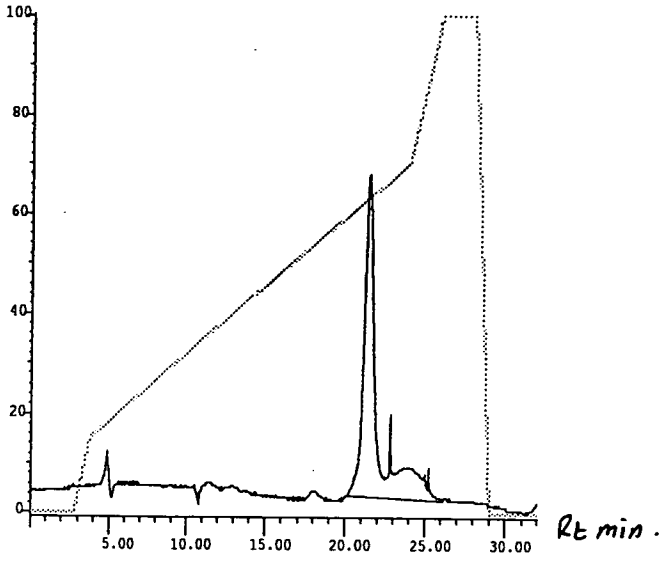
After exposure to various reaction conditions, eg. DBU / TEA / Water, Thiophenol / DBU and Thiophenol / TEA, the oligonucleotides were examined by HPLC and their retention times compared to a control sample of **6.1.** which had been purified as outlined in the experimental section. In the case of reactions involving oligonucleotide **6.2.** it was first of all necessary to deprotect and purify the sequence. This was carried out using the standard conditions described in the experimental section. This particular sequence was used since it is much easier to purify dodecamers, by gel filtration, than shorter sequences and also because the elution, on HPLC, of dC and T is earlier than 8MeOdG or 8OxodG.

Basic conditions were used for promoting the conversion of 8MeOdG to 8OxodG so it was necessary to establish that degradation of 8MeOdG would not result from the hot conc. NH_3 solution which is required for deprotection of the oligonucleotide. No degradation was observed on incubating both 8MeOdG and oligonucleotide **6.1.** in a solution of conc. NH_3 at 60°C over a 24 hour period.

Unfortunately none of the conditions used, ie. thiophenol / TEA or DBU and DMF / TEA / Water, showed any signs of converting 8MeOdG to 8OxodG, (Figures 3.5a and 3.5b). However, it was possible that oligonucleotides containing 8OxodG and 8MeOdG were co-eluting, and in order to establish whether or not any reaction had taken place it was necessary to subject the product oligonucleotides to enzyme digest and HPLC analysis such that the nature of the component nucleosides could be confirmed. Comparison of the elution profile of these digested samples against base standards indicated that all of the reaction conditions yielded an oligonucleotide containing the three components dC, T and 8MeOdG. This is clearly illustrated in Figures 3.5c and 3.5d.

Figure 3.5.

A) Elution profile of **6.1.**



B) Mixed injection. **6.1.** control and after reaction with DBU, TEA and water.

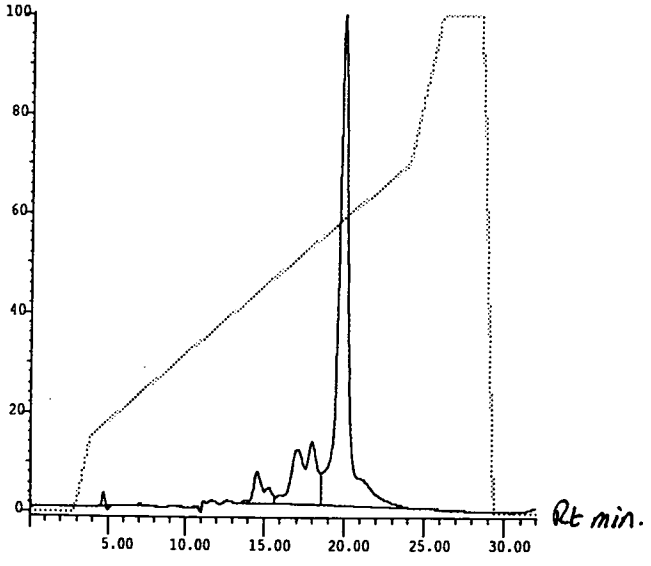
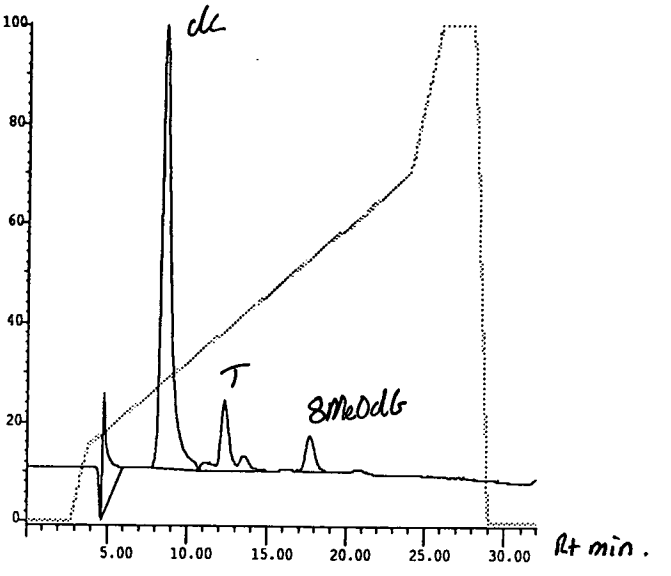


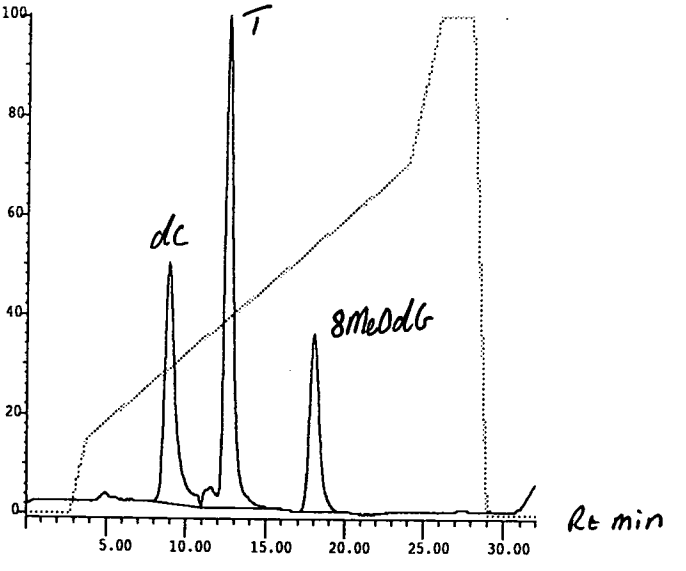
Figure 3.5.

C) Enzyme digest profile of **6.1.** after reaction with DBU, thiophenol and water.



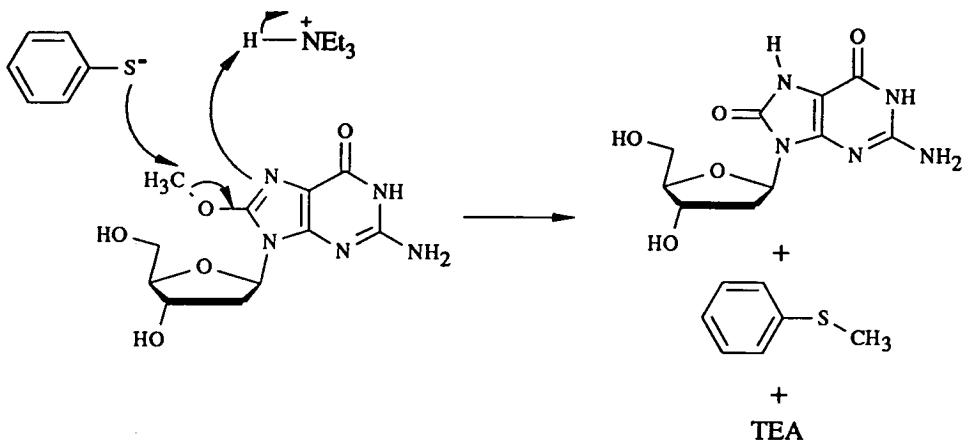
6·1 enzyme digest

D) Mixed injection of [illegible] with base standards, dC (8.8 min.), T (12.1 min.) and 8MeOdG (18 min.).



It is not clear why an 8MeOdG base contained in an oligonucleotide could not be converted into an 8OxodG residue. It is possible that the mode of action of thiophenol, in comparison to the same reaction at the nucleoside level, (Figure 3.6), is affected by the slightly higher pKa values of the ring nitrogens in the 5' and 3' phosphorylated nucleoside.

Figure 3.6.



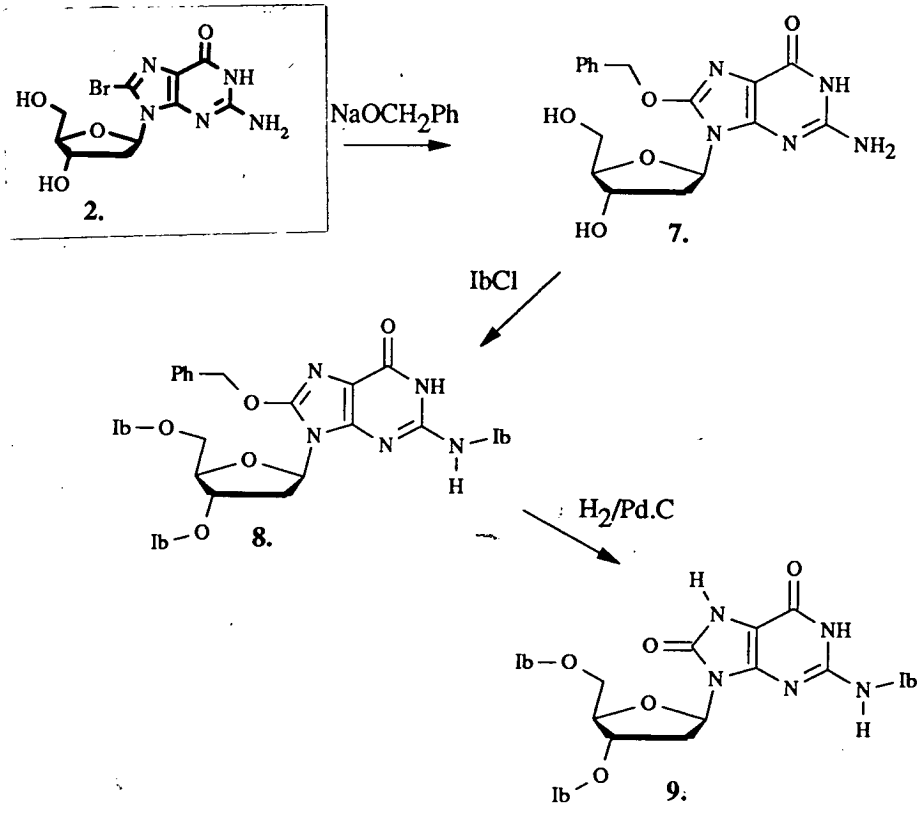
Conversion of oligonucleotides containing 8MeOdG to those containing 8OxodG was subsequently demonstrated, (Oda 1991), using a solution comprising thiophenol, TEA and DMF. Their choice of protection for the N^2 position of 8MeOdG was different from ours and our negative results may be a result of the slightly poorer electron withdrawing properties of the methylene-dimethylamine functionality in comparison to an amide moiety. The presence of an amide group might enhance the lability of the C-8 ether linkage to nucleophilic attack by lowering the electronegativity of the purine ring, the amidine group however, being less electron withdrawing, will not have same effect and will suppress nucleophilic displacement of the C(8)- Br. This would also explain the lack of reactivity of the unprotected oligoncleotide, **6.1.**

As a result of this lack of success in converting 8MeOdG to 8OxodG it was decided to investigate the alternative, Route B in Figure 3.1., involving the direct incorporation of 8OxodG into an oligonucleotide.

3.3. Synthesis of 8-BnodG nucleosides.

The synthesis of 8-benzyloxy-2'-deoxyguanosine (8BnOdG) **7.**, has been well documented and was straightforward, (Figure 3.7.). (TAI-SHUN, 1985).

Figure 3.7.



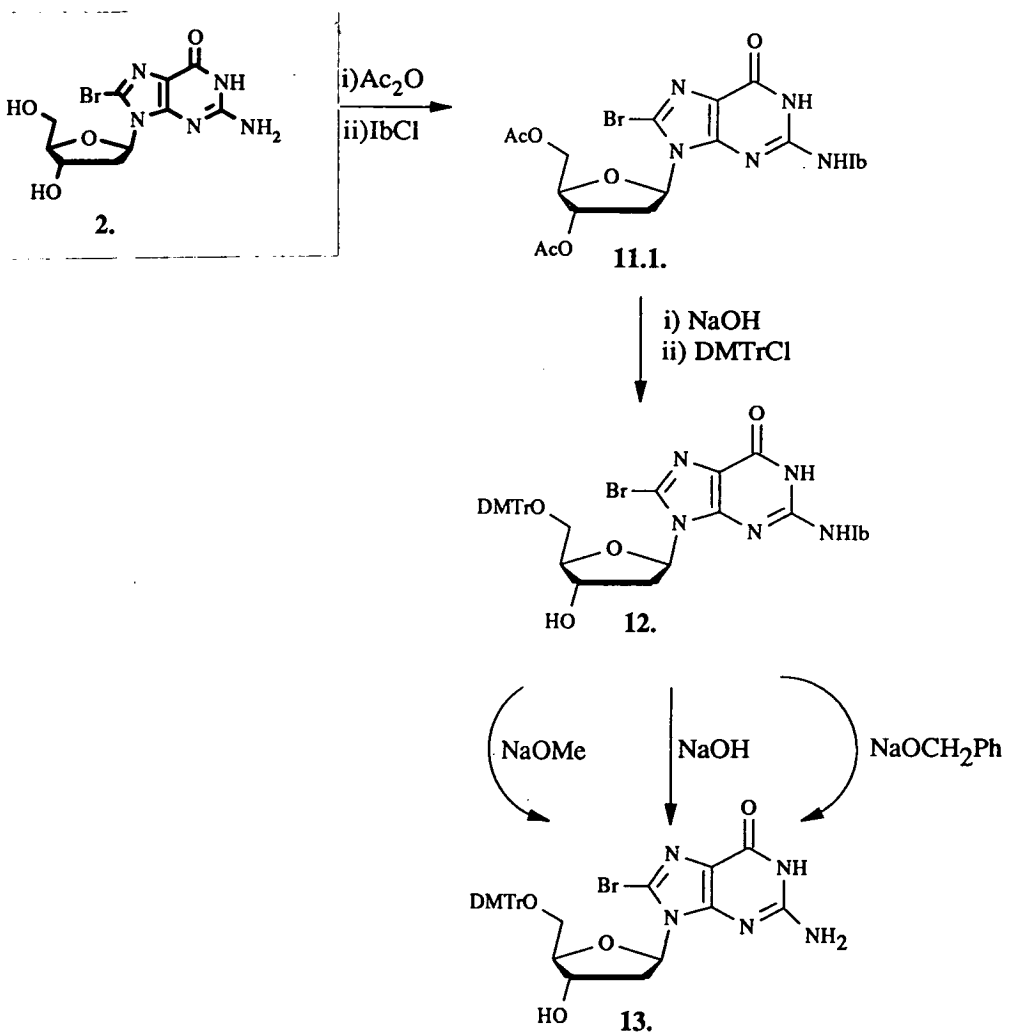
However, for the same reasons as for 8MeOdG, it was not possible to purify **7.** by crystallisation. Recrystallisation from methanol leads to only a 28% recovery, (Tai-Shun Lin 1985), so it was decided to convert **7.** to the triisobutyryl derivative, allowing purification of **8.** by silica gel chromatography, with subsequent hydrogenation yielding compound **9.**

However, due to the difficulty in producing large amounts of compound **7.** in a pure form, it was decided to use protected forms of 8BrdG to yield the 8-alkoxy derivatives. This would simplify the purification problem by rendering the products soluble in organic solvents, thus allowing the use of silica gel chromatography for purification at each stage of the synthesis. This would prevent impurities from one reaction being carried over into the next, a likely cause of the poor yields.

3.4. Reactions of 8BrdG.

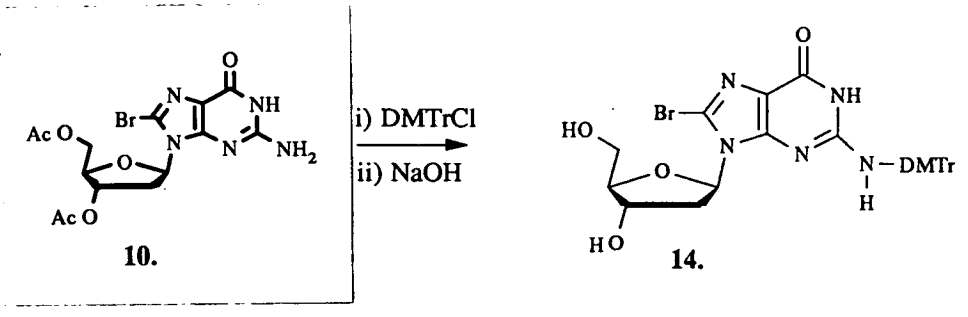
Therefore compound **12.** was prepared, (Figure 3.8.). Acetylation of the 8BrdG

Figure 3.8.



hydroxyl groups followed by attachment of an isobutyryl or phenylacetyl group yielded compounds **11.1.** and **11.2.** respectively. Subsequent saponification of the ester moieties followed by reaction with 4,4'-dimethoxytrityl-chloride afforded **12.** Reaction of **12.** with sodium hydroxide, benzyloxide or methoxide yielded only compound **13.** which was formed by simple cleavage of the base labile N^2 amide functionality. No displacement of the bromine was observed, (confirmed by FAB-MS: peaks at 416 and 418, in a 1:1 ratio). Therefore compound **14.** was prepared, (Figure 3.9.), which contains an N^2 acid labile dimethoxytrityl group.

Figure 3.9.



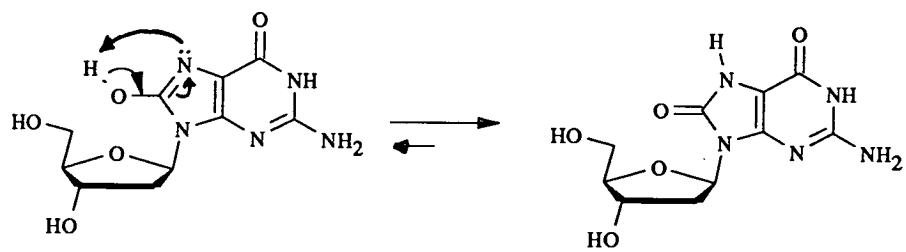
Attempts to displace bromine from this compound with hydroxide, methoxide and benzyloxide, with and without DMSO present, all proved unsuccessful with compound **14.** being extremely stable under all these conditions.

The lack of displacement of bromine from compounds **12.** and **14.** can be attributed to two effects, the large steric bulk of the dimethoxytrityl moiety and its net electron donating properties when attached to the N^2 position where it can directly interact with the purine ring resonance. This explains why 8BrdG derivatives, **12.** and **14.**, remain inactive to attack by the three anions, hydroxide, methoxide and benzyloxide, whereas the bromine atom of the unprotected nucleoside, compound **2.**, can be readily displaced by both methoxide and benzyloxide.

3.4.1. Reaction of 8BrdG with hydroxide.

A method for the preparation of alcohols is *via* the hydrolysis of the corresponding halogeno compound using sodium or potassium hydroxide (Morrison and Boyd). If it was possible to hydrolyse the C(8)-Br bond of 8BrdG, then it would react to form 8-hydroxydG which would rapidly tautomerise to the predominant 8-keto-7-hydro form, (8OxodG), (Figure 3.10.). Reaction of **2.** at 60°C showed no displacement of the bromine by hydroxide, and this can be readily explained by the very high pH of the reaction mixture. A strong base, rather than hydrolysing the C(8)-Br bond, will first of all abstract any base labile protons which are present.

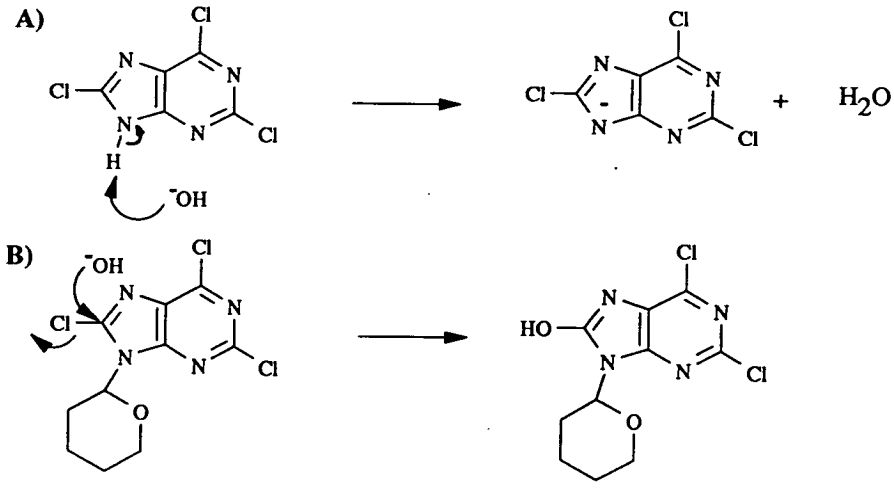
Figure 3.10.



8-hydroxydG. 8-OxodG.

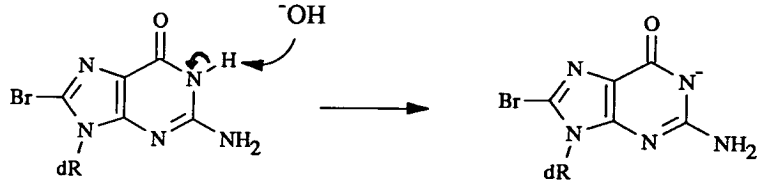
8BrdG has an N(1) proton with a pKa of 9.42 (Saenger 1984), therefore, when using strong bases such as sodium and potassium hydroxide the N(1) proton will be immediately abstracted resulting in an anionic purine ring. This will have the effect of inactivating the ring to substitution by nucleophilic reagents. Such inactivation has been well documented in the reaction of trichloropurine with hydroxide ions, (Figure 3.11a). If however the deprotonation at N(9) is prevented, by the attachment of a tetrahydro-2'-pyranyl group, (Figure 3.11b), then hydrolysis of the C(8)-Cl bond is observed, yielding the 8-hydroxy derivative (Suttcliffe 1963).

Figure 3.11.



For the same reasons, reaction of 8BrdG with strong base will not readily promote hydrolysis of the C(8)-Br bond since the N(1) position will be deprotonated, (Figure 3.12.).

Figure 3.12.

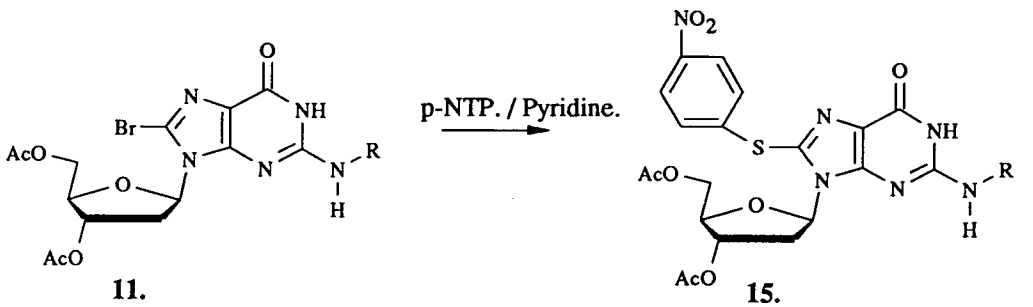


In order to assess the potential for bromine displacement, various nucleophiles, eg. azide, benzene sulphinic acid and various thiols, which can be activated by bases such as TEA, DBU and pyridine, were reacted with either compound **2.** or **10.** Acetate was also used as a nucleophile since it has been demonstrated (Guy 1988) that the treatment of 8BrdA with NaOAc in the presence of AcOH or Ac_2O leads to bromine displacement to afford the acetylated 8-Oxo derivative. This is not however a viable route to 8OxodG due to extensive depurination during the reaction. Indeed when this was examined, (Roelen 1991) it was found to result in a yield of 5', 3', N^2-

triacetyl-8OxodG of only 12%. 8-Oxo-7-hydro-guanosine can be readily prepared by this method since the depurination problems are less severe when dealing with ribosides as opposed to deoxyribosides, (Roelen 1991).

Only one nucleophile, p-nitrothiophenol, was found to displace bromine, and the reaction was carried out in pyridine at room temperature on compounds **11.1.** and **11.2.**, to produce the two derivatives **15.1.** and **15.2.** (Figure 3.13.).

Figure 3.13.



11. **15.**

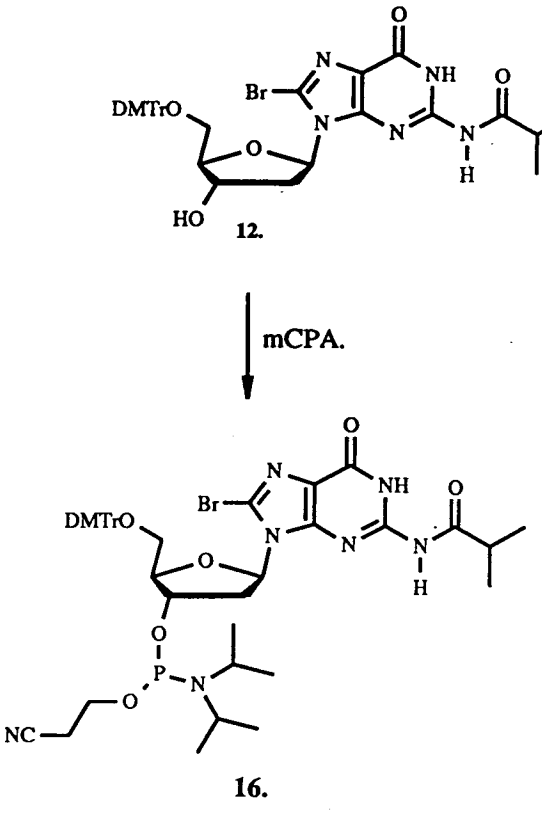
R = Isobutyryl, **11.1.** and **15.1.** R = Phenylacetyl, **11.2.** and **15.2.**

No other sulphur containing nucleophiles were found to displace the bromine and it was decided to incorporate 8BrdG into oligonucleotides to establish whether or not this would improve the lability of the C(8)-Br bond.

3.5. Incorporation of 8BrdG into oligonucleotides.

Synthesis of the 8BrdG phosphoramidite, **16.**, was carried out in accordance with the general procedure outlined by Gait,(Gait 1984). Reaction of **10.** with isobutyryl chloride afforded **11.1.** which was saponified with NaOH and reacted with 4,4'-dimethoxytrityl chloride to give **12.**, (Figure 3.8). Phosphitilation of the 3'-hydroxyl group then yielded the desired phosphoramidite **16.**, (Figure 3.14), which was incorporated, with 98.3% coupling efficiency, into oligonucleotides **17.** and **18a-e.** (see sect. 3.7).

Figure 3.14.



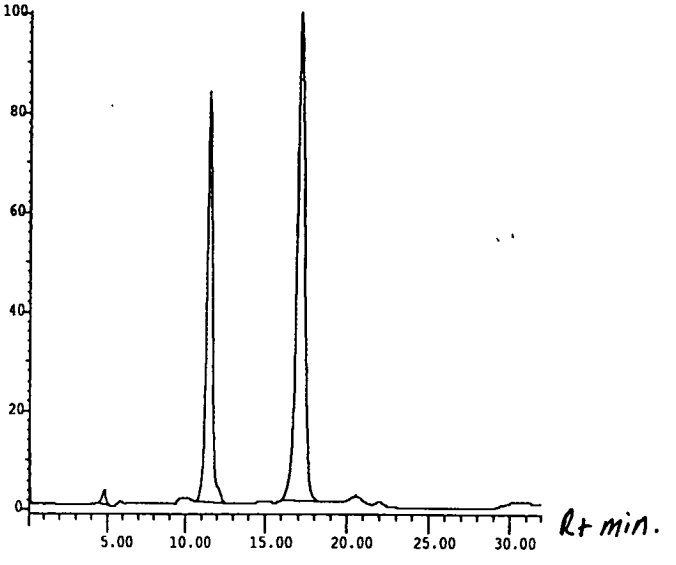
3.5.1. Reaction of d[TT8BrGTT] with nucleophilic reagents.

In order to fully assess the possibilities of displacing the bromine of 8BrdG with nucleophilic reagents, oligonucleotide **20.**, d[TT8BrGTT], was synthesised and subjected to various reagents. In all cases no displacement of bromine was observed. The product oligonucleotides all coeluted with the control oligonucleotide, d[TT8BrGTT], and HPLC analysis of the enzyme digest indicated only two components, T and 8BrdG, (Figures 3.15a and 3.15b).

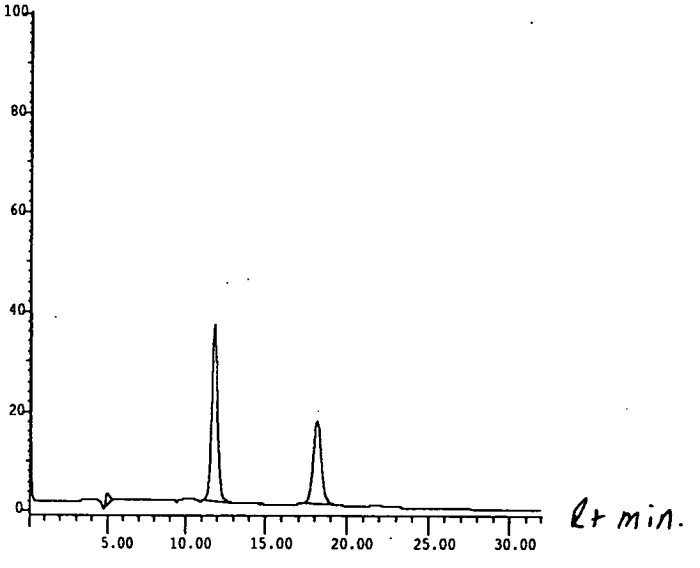
Figure 3.15.

A) Mixed injection, enzyme digest of **20.**, after reaction with p-nitrothiophenol, and standard 8BrdG + T.

Rt = 11.7 min. (T), Rt = 17.1 min. (8BrdG).



B) Standard 8BrdG and T.



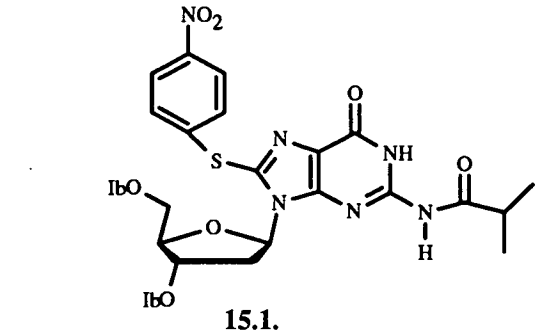
3.6. Displacement of bromine with p-nitrothiophenol.

As outlined in Figure 3.14., successful displacement of the bromine was achieved using p-nitrothiophenol, to generate compounds **15.1.** and **15.2.**, which was very important since it has been shown (Nikiforov 1991) that in certain cases p-nitrophenylthio, (NPT), can be displaced using potassium thioacetate to generate the thiol derivative. Furthermore, it has also been shown that conc. ammonia can be used to displace the p-nitrophenylthiol moiety to generate the keto derivative. Both these reactions were of great interest since both 8-oxo and 8-thio derivatives were potentially very useful. No substitutions were observed in our case however and disappointingly the 8NPTdG derivative was extremely stable to all conditions tested. It may be that severe steric crowding around the C(8) position makes these otherwise standard reactions very difficult due to the close proximity of the sugar moiety preventing attack from nucleophiles. Oxidation of the sulphur with mercuric chloride and iodine proved unsuccessful.

As a result of the stability of the compound to the reagents described above and also to hot conc. NH_3, it was a good candidate for incorporation into oligonucleotides where the presence of the C(8)-NPT moiety could perhaps be utilised for the binding of antibodies. A highly selective method whereby a specific functionality is bound by an antibody, after first of all creating an immune response, (Catty 1988). Before considering this method however, it is necessary to establish the effect of such a group upon the stability of the duplex. This can be determined by ultraviolet melting. In addition, it was necessary to establish whether or not the NPT functionality could be displaced by NH_3 or thioacetate when 8NPTdG was incorporated into an oligonucleotide. The synthesis of a suitable 8NPTdG phosphoramidite was carried out as illustrated in (Figure 3.16.). The phosphoramidite **21.** could have been synthesised using N^2-isobutyryl or N^2-phenylacetyl protection, but only the

isobutyryl derivative, **15.1.** was prepared which was suitably protected to yield monomer **21.** This was successfully introduced into oligonucleotides **22.a-d.**, see experimental section **4.3.**, with a coupling efficiency of 99.7%. No displacement of the NPT was observed.

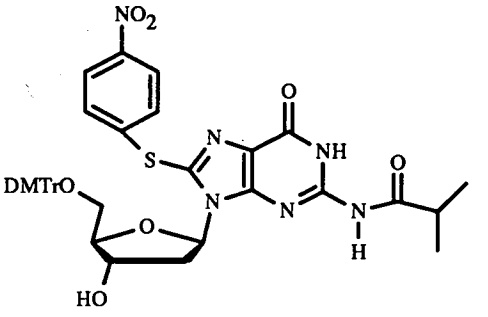
Figure 3.16.



15.1.

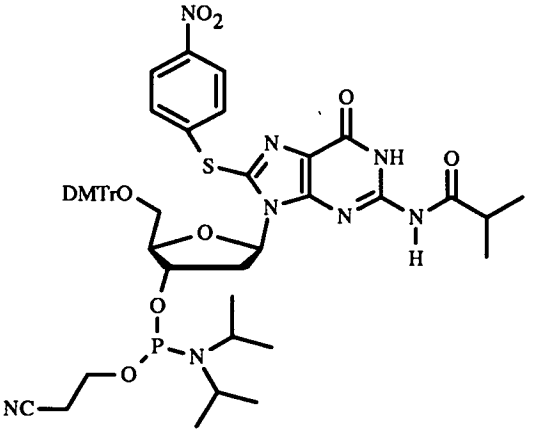
i)NaOH.

ii) DMTrCl.



20.

mCPA



21.

3.7. Thermodynamic stability of dodecamers containing C(8)- Br and NPT substituted dG.

Oligonucleotides **18a - e.** and **22a - d.** were examined by UV melting techniques such that the melting temperature, T_m, could be established enabling the effect of the substituents on the duplex stability to be assessed.

The melting temperatures, (Table 3.1), demonstrate the destabilisation of the duplex that occurs when either bromine or p-nitrophenylthio are incorporated at the C-8 position of dG. The presence of two 8NPTdG.dC base pairs, the most stable base pair in this series of oligonucleotides, resulted in an oligonucleotide with a T_m which was 28.2°K lower than that for the native dG.dC oligonucleotide.

Table 3.1.

Oligonucleotide.	**T_m°K.**	
d[CGCGAATTCGCG]	338.2	
	X=8BrdG.	**X=8NPTdG.**
d[CGCXAATTAGCG]	283.2 (18a)	---------
d[CGCXAATTGGCG]	322 (18b)	304.2 (22a)
d[CGCXAATTCGCG]	330 (18c)	310 (22b)
d[CGCXAATTTGCG]	310.8 (18d)	283.4 (22c)
d[CGCAAATTXGCG]	301.5 (18e)	296.1 (22d)

All other base pairs containing 8NPTdG, yielded duplexes of extremely low stability. This can possibly be attributed to the bulkiness of the group which creates severe steric hindrance and inhibits duplex formation. As a consequence of this unstable duplex, 8NPTdG is unsuitable for incorporation into oligonucleotides since there is

severe disruption of the normal pairing and stacking interactions necessary for formation of stable helical DNA.

When 8BrdG was incorporated into oligonucleotides the resultant destabilisation was much less dramatic, indicating the ability of the duplex to accommodate the bromine atom. Oligonucleotides **18b.** and **18c.** are the most stable of the series with T_m values 16.2 and 8.2°K less than the native dG.dC dodecamer.

Although this work does not yield any direct evidence on the pairing of 8BrdG with the four naturally occuring nucleotides, it does however demonstrate that the incorporation of a C(8)- bromine substituent will not drastically alter the base pairing properties, and the ability of the oligonucleotide to form a nearly normal helix. The presence of such C(8) substitutions can have the effect of altering the base into a *syn* orientation, (Uesugi 1977), and although it is not possible to establish whether or not this occurs in the above sequences, this effect may have consequences when 8BrdG is incorporated into dG rich sequences.

Oligonucleotides containing dG rich regions are notoriously difficult to purify due to unfavourable secondary structure formation. Secondary structures involving dG are well known to occur in telomeric regions, dG rich regions at the end of chromosomes, which help to seal the chromosome and prevent fraying of the ends (Chul-Hee 1992). The structure of telomeres are known to involve dG tetrads which utilise both the Hoogsteen and Watson / Crick hydrogen bonding sites and rely on alternate dG's adopting a *syn* orientation. As a result, it was decided to incorporate 8BrdG into such sequences at every alternate dG in the dG rich region to determine whether or not the presence of such a base would interfere and destabilise these undesirable structures. Its incorporation did not however prevent the formation of these secondary strucure and as a result did not resolve the purification problems, indicating that 8BrdG was probably accommodated into the duplex in the same

manner as dG. For another reason however, this has proved to be an extremely beneficial property since the presence of a bromine atom, with its high electron density, can be used as a heavy atom marker in aiding the refinement of oligonucleotide X-ray crystallographic data.

Difficulty in solving the crystal structure of a d[TGTACA] / nogalomycin complex, has highlighted the need for the incorporation of a nucleotide containing a heavy atom marker. As such, 8BrdG was used to replaced dG in the above sequence and the resultant d[T**8BrG**TACA] / nogalomycin complex was crystallised in the same space group as the native complex, further demonstrating that the C(8)-Br does not unduly interfere with formation of the helix. Thus the X-ray data, for the complex containing 8BrdG, which has been collected to 2.5 *A*, can be used to aid the structure solution.

3.8. Conclusion.

Incorporation of 8MeOdG into oligonucleotides was carried out very successfully, although the synthesis of the 8oxo derivative was not possible. 8OxodG has however been utilised in oligonucleotides, (Kouchakdjian 1991 and Oda 1991). Kouchakdjian synthesised an 8OxodG phosphoramidite *via* the 8-benzyloxy intermediate which was readily hydrogenated to 8OxodG. The methodology adopted by Oda was to synthesise oligonucleotides containing an 8MeOdG nucleotide containing acetyl protection on the N^2 amine. The methoxy moiety was then converted into the oxo derivative by action of thiophenol prior to deprotection of the oligonucleotide. Our oligonucleotides contained 8MeOdG nucleotides with dimethylaminomethylene protection on the N^2 amine. By using this form of protecting group it allowed us to selectively protect the N^2 amine and react the product with dimethoxytrityl chloride without purification of the intermediate. In contrast, amide protection, eg. isobutyryl, yields the 5', 3'-(O)-N^2- tris isobutyryl derivative which requires removal of the ester

moieties, prior to reaction with dimethoxytritylchloride, and we found this extremely problematical when using both acetyl and isobutyryl as the amide protection.

Recent work however, (Bodepudi 1992), has indicated the best method for synthesis of an 8OxodG nucleoside. The work described has significantly improved the yields of 8BrdG and 8BnOdG to 84%, previously 55%, and 70%, previously 28%, respectively. Protection of the amine with the isobutyryl moiety followed by hydrogenation yielded the desired 8-oxo derivative which was easily transformed into the corresponding nucleoside phosphoramidite and incorporated into DNA with a > 98% coupling efficiency.

Another alternative, for more diverse C(8) modification, would involve the synthesis of an O^6 protected 8BrdG. This form of protection would have the effect of deprotonating the N(1) position thus allowing more facile displacement of the bromine in basic conditions. It is conceivable that a suitable group, such as mesitylene sulphonyl, (Li 1989), or 4-nitrophenethyl, (Gaffney 1982), would be sufficiently stable to enable displacement of the bromine by anions such as azide, acetate or hydroxide allowing the synthesis of a variety of C(8) substituted dG derivatives. Such substituted nucleotides would find use in areas such as DNA / protein crosslinking, protein / DNA recognition and in the study of mismatched base pairs.

4. Experimental section.

4.1. Organic solvents.

DNA synthesis grade acetonitrile and all other reagents required for the synthesis of oligonucleotides were purchased from Applied Biosystems Ltd. HPLC grade acetonitrile was purchased from Fisons and anhydrous dimethylformamide (HPLC grade) was purchased from Aldrich.

Purification of organic solvents.

Methanol was dried by distillation from magnesium turnings and iodine; Dichloromethane and Pyridine were dried by distillation from CaH_2. Dimethyl sulphoxide was dried by stirring overnight with CaH_2 followed by distillation under reduced pressure (b.pt. 76°C / 1.6 x 10^3 Nm^{-2}). Tetrahydrofuran was dried by distillation from sodium benzophenone. Benzyl alcohol was purified by distillation under reduced pressure (b.pt. 65°C / 750 Nm^{-2}). Diethyl ether and toluene were dried by addition of sodium wire. N,N'-diisopropylethylamine was dried by storage over CaH_2.

4.2 Experimental Techniques.

4.2.1. Thin layer chromatography (t.l.c.).

All t.l.c. analysis was carried out using Merck aluminium sheets coated with silica gel 60 F_{254}. After elution, the chromatograms were examined under ultraviolet light. For use with molecules containing the acid labile 4,4'-dimethoxytrityl moiety, the silica plates were soaked in triethylamine and dried prior to use. Visualisation of components present on the tlc. plate was carried out using the reagents described below.

Visualisation of t.l.c. chromatograms.

Presence of sugars: A solution comprising anisaldehyde (5 ml), acetic acid (1 ml), sulphuric acid (1 ml) and ethanol (100 ml), was sprayed onto the t.l.c. plate which was then heated on a hotplate. A positive test was indicated by a blue / green colouration on heating. This test also allowed positive identification of compounds containing the 4,4'-dimethoxytrityl group due to the production of an orange colour immediately upon spraying.

Presence of the 4,4'-dimethoxytrityl moiety: Holding the t.l.c. plate over a beaker of conc. hydrochloric acid produced an intense orange colour.

Presence of the 2-cyanoethyl-N,N'-diisopropyl-phosphoramidite moiety: A 0.1% solution of dithio-*bis* p-nitrobenzoic acid in ethanol / Tris buffer (0.45 *M*, pH 8.2), 1 : 1, (Ellmans reagent), was sprayed onto the t.l.c. plate to produce an instantaneous bright yellow colour.

Solvent systems for t.l.c. (all v/v).

A. Dichloromethane (4), methanol (1), acetic acid (10 drops).

B. Dichloromethane (6), methanol (1), acetic acid (10 drops).

C. Dichloromethane (9), methanol (1), acetic acid (10 drops).

D. Dichloromethane (4), methanol (1).

E. Dichloromethane (7), methanol (1).

F. Dichloromethane (9), methanol (1).

G. Dichloromethane (95), methanol (5).

H. n-Butanol (3), water (1), acetic acid (10 drops).

I. i-Propanol (3), water (1), ammonia (10 drops).

J. Ethanol (8), water (2).

K. Ethylacetate (5), dichloromethane (5).

L. Ethylacetate (8), dichloromethane (2), TEA (10 drops).

M. Dichloromethane (95), methanol (5), TEA (10 drops).

N. Acetonitrile (95), ethylacetate (5) and TEA (10 drops).

O. Acetonitrile (2), dichloromethane (8).

4.2.2. Oligonucleotide crystallisations.

The oligonucleotide (0.9 mg) was dissolved in sodium cacodylate (50 m*M*, 45 µl, pH 7.0) and 5 µl of this solution was placed into the first well of a nine well crystallising plate. Specific quantities of solutions A, B and C were then added sequentially, (see below), in that order, ensuring that the oligonucleotide did not precipitate.

Solution A: $MgCl_2$ (100 m*M*).

Solution B: 50% aqueous 2-methyl-2,4-pentanediol.

Solution C: Spermine.4HCl (10 m*M*).

This protocol was then applied to each of the remaining eight wells, with the amount of solutions A - C being varied slightly in each well, by 1 or 2 µl, in order to establish the best conditions for crystal formation. The wells were then covered with glass microscope slides with vacuum grease applied to ensure an efficient seal. The plate was maintained at 277°K and examined daily under a microscope to follow the process of crystallisation.

d[CGCO^{6}MeGAATTTGCG].

The self-complementary dodecamer d[CGCO^{6}MeGAATTTGCG]$_2$ is closely related in sequence to the native dodecamer (Wing 1980). Crystals were grown at 277°K and pH 6.3 from 25 µl drops containing the oligonucleotide (1.0 mM), magnesium chloride (33 mM), 2-methyl-2,4-pentanediol (16.7%), spermine (1.25 mM) and sodium cacodylate (5 mM). A single crystal of dimensions 1.3 x 0.3 x 0.3 mm was

mounted in a sealed glass capillary and X-ray data collected by Dr. G. Leonard, as discussed in appendix i.

4.2.3. Nuclear Magnetic Resonance.

^{1}H NMR spectra were recorded on the following spectrometers.

Brucker WH360, Brucker WP200 and Brucker WP80. All chemical shifts were quoted in ppm relative to trimethylsilane ($\delta = 0$).

^{31}P NMR were recorded on a Jeol JMN FX90Q spectrometer.

4.2.4. Mass Spectrometry.

Mass spectra were obtained by positive ion Fast Atom Bombardment (FAB) on a Kratos MS50 mass spectrometer in a matrix of either thioglycerol or 3-NOBA.

4.2.5. Elemental analysis.

C, H and N micro analysis was performed on a Perkin Elmer 240 elemental analyser.

4.2.6. Ultraviolet spectra.

All ultraviolet spectroscopy was carried out using a Perkin Elmer Lambda 15 spectrophotometer controlled by an IBM-PS2 microcomputer equipped with the PECSS software package.

4.2.7. Melting points.

These were obtained using a Kofler Hotstage and are uncorrected.

4.2.8. Flash Chromatography.

This was carried out under slight nitrogen pressure (0.5 bar) using Fluka silica 60,

mesh (230-240). In all cases the silica gel was suspended in the eluent and sonicated for 15 min. prior to wet packing.

Technique A:- For compounds containing the 2-cyanoethyl-N,N'-diisopropyl-phosphoramidite functionality the silica was oven dried at 120°C, cooled in a desiccator suspended in eluent containing 2% triethylamine and sonicated for 15 min. After wet packing the column was washed with three column volumes of eluent before loading the sample.

4.2.9. Oligonucleotide synthesis and purification.

All oligonucleotides were synthesised by the solid phase method on an Applied Biosystems 380B or 381A DNA synthesiser, using standard phosphoramidite chemistry (Beaucage 1992 and 1981, Matteucci 1981 and Letsinger 1976). Synthesis reagents were purchased from Applied Biosystems except for 5'-(4,4'-dimethoxy-trityl)-N^2-isobutyryl-O^6-methyl-2'-deoxyguanosine-3'-[2-cyanoethyl-N,N'-diisopropyl]-phosphoramidite, which was purchased from Glen Research, 44901 Falcon Place Sterling VA. 22170

Oligonucleotides were cleaved from the solid support with concentrated aqueous ammonia, deprotected by heating the solution at 60°C for 8 hours and subsequently purified by anion-exchange chromatography (DEAE cellulose), followed by reversed phase HPLC (method 2). Desalting of the oligonucleotide was carried out on a Sephadex G-10 column prior to lyophilisation. For oligonucleotides containing O^6MedG, the following protocol was observed. The 5'-dimethoxytrityl-N^2-isobutyryl-O^6-methyl deoxyguanosine-3'-cyanoethyl phosphoramidite monomer was used to introduce O^6MedG and the fully assembled oligonucleotide was cleaved from the solid support and deprotected in a 5% solution of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in anhydrous methanol for two weeks at ambient temperature in an atmosphere of nitrogen (Kuzmich 1983). At no time was

the oligonucleotide exposed to ammonia, as this can lead to the slow conversion of O^6MedG to 2,6-diaminopurine (Borowy-Borowski 1987). The DBU was removed with Dowex-50 cation exchange resin (Na^+ form). Subsequent purification and lyophilisation was carried out as above.

4.2.9a. Determination of phosphoramite coupling efficiency.

During oligonucleotide synthesis the 4,4',-dimethoxytrityl moiety is cleaved from the growing chain using a dichloromethane solution, containing 3% trichloroacetic acid, prior to the addition of the next monomer unit. The removal of this protecting group generates a 4,4'-dimethoxytrityl cation which has a vivid orange colour. Collection of these trityl fractions after each monomer addition and measurement of the visible absorbance at 495 nm, (A_{495}), of each fraction gives an indication of the coupling efficiency. This was determined as shown below.

Each fraction was diluted with toluene-4-sulphonic acid (0.1 *M*) in anhydrous acetonitrile to a volume of 100 ml into a standard flask and the absorbance at 495 nm determined spectrophotometrically, the coupling efficiency was established according to the following equation.

$$\mathbf{C.E. = X / Y \times 100\%}$$

Where :-

C.E = Coupling Efficiency.

X = A_{495} of trityl fraction n + 1.

Y = A_{495} of trityl fraction n.

4.2.9b. Determination of oligonucleotide concentration.

Oligonucleotides were synthesised in multimilligram quantities by combining several 1 μM syntheses. The concentration of oligonucleotides however, is best defined in terms of optical density per millilitre (OD / ml), an arbitrary unit, which corresponds to the UV absorbance of a 1 ml aqueous solution of the oligonucleotide at 264 nm, in a 1 cm path length. The optical density is determined by the following equation:

$$OD = (A_{264} \times V_c \times V_{tot.})/V_a$$

Where:

V_a = Volume of the oligonucleotide aliquot added to the UV cell.

V_c = Volume of water contained in the UV cell.

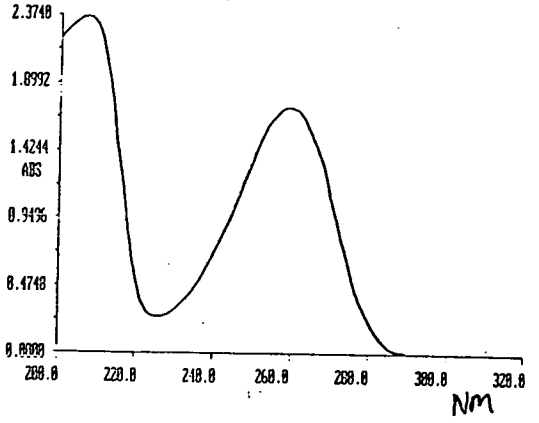
$V_{tot.}$ = Total volume of oligonucleotide sample.

4.2.9c. Reversed Phase HPLC.

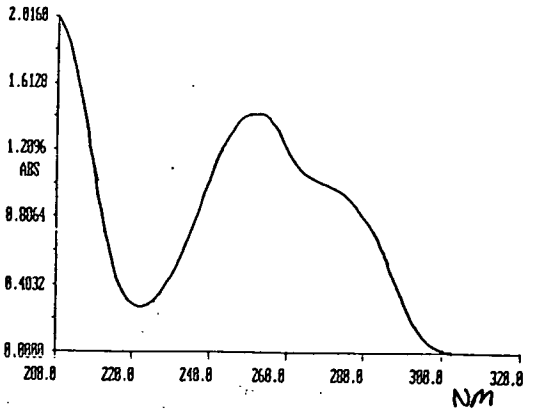
The bases ~~involved in~~ oligonucleotide chains all have characteristic UV spectra with maxima in the region 250 to 300 nm, (Figures 4.1a to 4.1e). Hence, the HPLC elution profile of an oligonucleotide is monitored between these limits, usually at 264 nm for a dilute sample (analytical injection) and 294 nm for a more concentrated preparative injection. As can be seen from the UV spectrum of a typical oligonucleotide, Figure 4.1f [dCGCO^{6}MeGAATTTGCG], there is a broad maximum ca. 257 nm which drops off to zero by 305 nm. Therefore careful monitoring around 260 nm allows the detection of very small amounts of oligonucleotide impurities and by "de-tuning" and adjusting the wavelength to 294 nm purification of larger samples can be acheived without overloading the detector.

Figure 4.1.

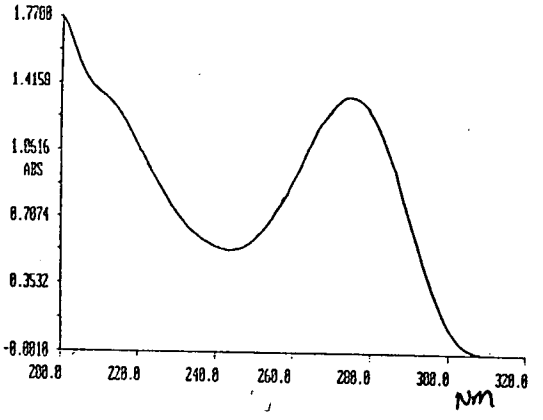
UV Spectra of nucleotide bases.



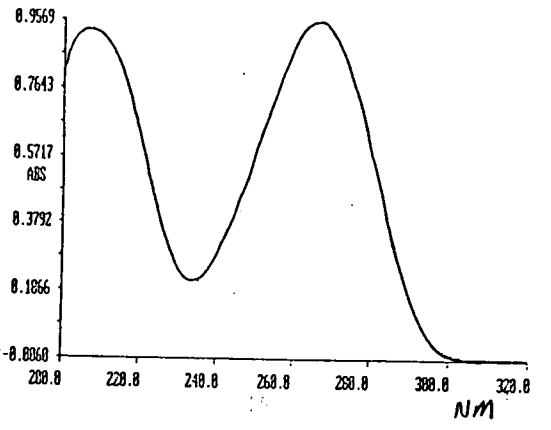
a) deoxyadenosine.



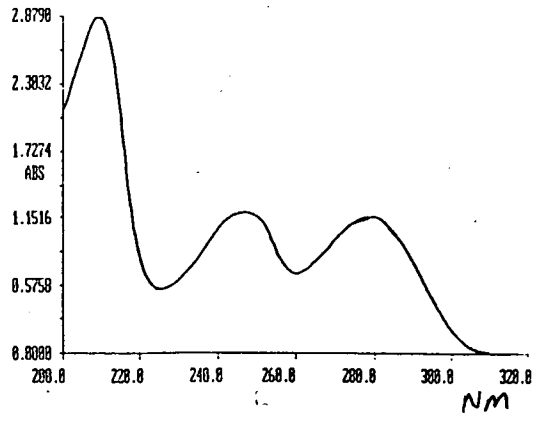
b) deoxyguanosine.



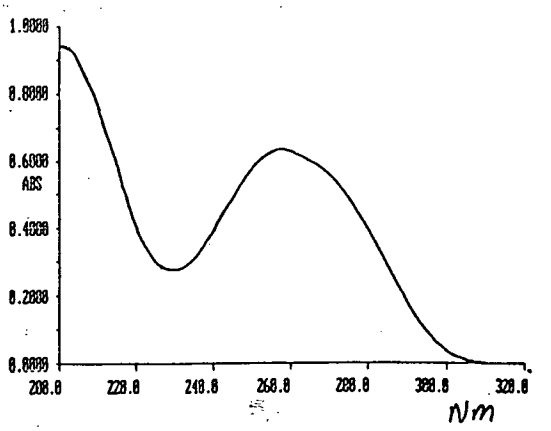
c) deoxycytidine.



d) Thymine.



e) O^6Methyldeoxyguanosine.



f) Oligonucleotide. d[CGCGO^6MeGAATTTGCG]

Method 1. Method employed for the analysis of enzyme digests of oligonucleotides containing O^6MedG.

Detection: UV absorbance at 264 nm.

Buffer A: Potassium acetate (10 m*M* adjusted to pH 5.0 with TFA).

Buffer B: Acetonitrile : Water (3 : 1, v/v).

Conditions

Flow rate = 2 ml / min.

Time(min)	0	6	11	15	16	26
Buffer A(%)	95	95	70	70	95	95
Buffer B(%)	5	5	30	30	5	5

Method 2. Method employed for the purification of oligonucleotides and enzyme digest analysis.

Detection: UV absorbance. Analytical at 264 nm, Preparative at 294 nm.

Buffer A: Ammonium acetate (100 m*M*, pH 7.3).

Buffer B: 20% Acetonitrile in aqueous ammonium acetate (100 m*M*, pH 7.3).

Conditions

Flow rate = 3 ml / min.

Time(min)	0	3	4	24	25	28	29	30
BufferA(%)	100	100	90	30	0	0	100	10
BufferB(%)	0	0	10	70	100	100	0	0

Method 3. This method was used to establish the identity of 8OxodG by coelution with a standard sample. The work was carried out by Dr. E. Booth at the Shell

Research Centre, Sittingbourne, using Knaver 64 (2) pumps connected to an Ultrasphere ODS (10 mm x 250 mm) column controlled by a Kontron series 2000 programmer. Sample detection was obtained using a Rapiscan SA 6508 UV detector with elution being carried out isocratically using (8%) methanol / water at a flow rate of 2.5 ml / min.

HPLC purification of oligonucleotides.

All HPLC analyses were carried out on a Perkin Elmer series 410 LC HPLC system or with a Gilson system 306 equipped with a 115 UV detector and an 811B dynamic mixer controlled by a DCS computer with Gilson 712 software. Each system was equipped with an Aquapore, Octyl (C8) column, and an octadecyl (C18) Pellicular guard column.

Analytical HPLC ws carried out using method 2 above by injecting a solution of the oligonucleotide (0.1 OD) in ammonium acetate (100 m*M*, 0.7 ml), and monitoring at 264 nm and 0.1 absorbance units full scale (AUFS).

Preparative HPLC was carried out using method 2 and by injecting the product of a 1 μM DNA synthesis (50 OD approx.) dissolved in ammonium acetate (100 mM, 0.7 ml) and monitoring at 295 nm and 1 AUFS.

4.2.10. Enzymatic digestion of oligonucleotides.

Oligonucleotide digestion by enzymatic methods allows the identification of the nucleoside composition. The two enzymes used in this procedure are snake venom phosphodiesterase-1, which cleaves the oligonucleotide into its constituent nucleoside-5'-phosphates, and alkaline phosphatase, which catalysis the hydrolysis of the nucleoside-5'-phosphates into nucleosides.

Solution A: - Phosphodiesterase-1 (0.5 mg, 0.5 units) was placed in an eppendorf

tube and dissolved in water (50 µl).

Phosphodiesterase-1 (Type vii: Crotalus Atrox venom, 0.1 unit/mg of solid). 1 unit will hydrolyse 1.0 mM of *bis*(p-nitrophenyl)phosphate per minute (pH 8.8, 37°C)

Solution B:- Alkaline Phosphatase (1 µl, 0.15 units) was placed in an eppendorf tube and dissolved in water (19 µl).

Alkaline Phosphatase (Type iii-N:*E-Coli*, 31 units/mg of protein). 1 unit will hydrolyse 1 mM of p-nitrophenylphosphate (pH 10.4, 37°C).

Method.

The oligonucleotide (1 OD) was dissolved in buffer [NaCl (1*M*), Tris (50 m*M*, pH 8.8 with HCl (1*M*)]. Phosphodiesterase (solution A, 5 µl) and alkaline phosphatase (solution B, 1 µl) were added and the solution maintained at 37°C overnight before being examined by RP-HPLC, method 1, (injecting 0.1 OD of the sample).

4.2.11. Ultraviolet melting studies.

All data were recorded on a Lambda 15 ultraviolet-visible spectrophotometer controlled by an IBM-PS2 microcomputer. Ultraviolet melting data were collected and processed using the PECSS software package. The temperature of the sample cell was controlled by a Peltier block with a heating rate of one degree centigrade per minute. The absorbance (264 nm) of the sample was measured at intervals of ten seconds and the data were converted from absorbance versus time to absorbance versus temperature by the PECSS software. The melting temperature (T_m) was defined as the temperature which corresponds to the maximum point of the first derivative of the melting curve. All curves were determined in triplicate.

Theory of Ultraviolet melting (relating T_m to C_t).

The ultraviolet melting technique depends upon the assumption that two complementary single strands of DNA are in equilibrium with the double stranded form and that the double stranded form has a lower UV absorbance than single strands. This difference in absorbance between the two forms is known as the hyperchromicity of DNA. Therefore if the initial DNA single strand concentration (C_t) is known, then by relating this with the melting temperature, (T_m), thermodynamic parameters can be derived as shown below in Equation 1. (Marky 1987)

Equation-1.

For self-complementary oligonucleotides

$$A + A \rightleftharpoons B$$

1) Total concentration of A is C_t.

2) Total concentration of B is $C_t/2$.

3) If α is the fraction of strands forming a duplex at any given time then we can write the concentrations of A and B at any given time as:

$[A] = (1-\alpha)C_t$ and $[B]=\alpha C_t/2$

$K = [B]^b/[A]^a$

$K_{self.comp.} = \alpha/(1-\alpha)^2.2C_t$

At T_m $\alpha = 0.414$, $K = 0.602/C_t$

$K_{self.comp.} = 1.204/C_t$

$-RTlnK = \Delta H\phi - T\Delta S\phi$

$\ln 1.204/C_t = -\Delta H\phi/RT_m + \Delta S\phi/R$

$$\ln C_t = \Delta H\phi/RT_m - \Delta S\phi/(R + \ln 1.204)$$

$$\mathbf{1/T_m = R.\ln C_t/\ \Delta H\phi + [\Delta S\phi - \ln (1.204).R]/\ \Delta H\phi}$$

From equation 1 it is clear that $T_m \propto \ln C_t$ and therefore a plot of $1/T_m$ against $\ln C_t$ will yield a straight line of gradient $R/\Delta H\phi$ which bisects the Y-axis at $[\Delta S\phi - \ln(1.204).R]/\Delta H\phi$.

4.2.11a. Measurement of T_m as a function of C_t.

Using 5 mm path length UV cells, the spectrophotometer was background corrected with buffer in both cells [NaCl (1 *M*), NaH_2PO_4 (10 m*M*), EDTA (1 m*M*), pH 7, with NaOH (1 *M*)]. To the sample cell, containing buffer (1.4 ml), was added the oligonucleotide (2.8 OD) and the data was collected as described above. Successive oligonucleotide dilutions were then made, and the T_m was determined at six different oligonucleotide concentrations over a 20 fold range. The last two dilutions were made by transferring the oligonucleotide solution into a 1 cm path length cell.

4.2.11b. Measurement of T_m as a function of pH.

The sample cell was firstly filled with buffer [NaH_2PO_4 (0.1 *M*), adjusted to the appropriate pH by addition of NaOH (1 *M*)] and a baseline correction applied against a reference cell containing the same buffer. The oligonucleotide (1.2 OD) was then added to the sample cell and the data was collected as described above.

4.3 Synthetic chemistry.

1.1a. *Oligonucleotide* d[CGCQ^{6}MeGAATTCGCG].

1.1b. *Oligonucleotide* d[CGCQ^{6}MeGAATTTGCG].

2. *8-Bromo-2'-deoxyguanosine.*- To a rapidly stirred suspension of 2'-deoxyguanosine (5g, 14.4 mM) in water (30 ml) was added, dropwise, saturated bromine water, (*c.a.* 150 ml), until the yellow colouration persisted. The mixture was then cooled in an ice / water bath and examined by (t.l.c. system H. Rf_{sm} = 0.4, $Rf_{prod.}$ = 0.6). Saturated bromine water was then added in aliquots until no 2'-deoxyguanosine remained. The cold suspension was then filtered, washed with cold water followed by cold acetone and recrystallised from water / methanol (30:1, 1.5 L), maintaining the temperature below 80°C, to give 8-bromo-2'-deoxyguanosine (3.77g, 58%) as white needles. m.pt. 207°C dec. (Holmes 1964, 210°C dec.). [Found 34.4%C, 3.4%H and 20.2%N, Calc. 34.7%C, 3.5%H and 20.2%N]. δ_H (360 MHz; d^{6}DMSO) 2.1 (qd, 1H, 2'H), 3.1 (quin. 1H, 2'H), 3.4 (quin. 1H, 5'H), 3.55 (quin. 1H, 5'H), 3.8 (sext. 1H, 4'H), 4.2 (m, 1H, 3'H), 4.8 (t, D_2O exch. 1H, 5'OH), 5.2 (d, D_2O exch. 1H, 3'OH), 6.1 (t, 1H, 1'H), 6.3 (s, D_2O exch. 2H, N^{2}H), 10.8 (s, D_2O exch. 1H, N(1)H). m/z 345 and 347 $(M+H)^+$. λ_{max} (pH 5.6) 260 nm, 278(sh).

3. *8-Methoxy-2'-deoxyguanosine.*- To a stirred solution of sodium methoxide, 0.8 g Na in methanol 15 ml, was added anhydrous DMSO (20 ml) followed by a solution of 8-bromo-2'-deoxyguanosine (2g, 5.8 mM) in anhydrous DMSO (40 ml). The reaction mixture was protected from moisture and heated at 60°C for 8h. then cooled to room temperature and neutralised with Dowex (NH_4^+ form). After filtering the Dowex the clear filtrate was concentrated under reduced pressure and precipitated into rapidly stirred diethyl ether (1 L) to yield the product as a pale yellow solid. (H. Rf = 0.55). δ_H (200 MHz; d^{6}DMSO) 2.1 (br. m, 1H, 2'H), 2.85 (br. m, 1H, 2'H),

3.25 to 3.6 (br. m, 2H, 5'H), 3.8 (m, 1H, 4'H), 3.95 (s, 3H, C8-OMe), 4.35 (m, 1H, 3'H), 4.75 (br. s, 1H, D_2O exch. 5'OH), 5.25 (br. s, 1H, D_2O exch. 3'OH), 6.1 (t, 1H, 1'H), 6.75 (s, D_2O exch. 2H, N^2H), 10.7 (br. s, D_2O exch. 1H, N(1)H). [Found: m/z = 298.11512 $(M+H)^+$ Calc. for $C_{11}H_{16}N_5O_5$ = 298.11513] m/z* 298 $(M+H)^+$ and 181 $(M+H\text{-}dR)^+$. λ_{max}(pH 5.6) 282nm, 247nm, (pH 14) 260nm, 270nm(sh).

*Obtained by laser desorption time of flight mass spectrometry.

4. *5'-O-(4,4'-Dimethoxytrityl)-N^2-[(dimethylamino)methylene]-8-methoxy-2'-deoxyguanosine.*- 8-Methoxy-2'-deoxyguanosine (1.1g, 3.7mM), dried by evaporation of anhydrous pyridine (3 x 10 ml), was suspended in anhydrous pyridine (15 ml), protected from moisture and dimethylformamide diethyl acetal (2.2 ml, 4 eq) was added. The mixture was stirred overnight at 40°C, concentrated *in vacuo* and the oily residue was dried by evaporation of anhydrous pyridine (3 x 10 ml) before being dissolved in anhydrous pyridine (15 ml) and protected from moisture with a guard column. To this solution was then added, with stirring, triethylamine (1.1 ml, 2 eq), 4-dimethylaminopyridine (22 mg, 0.05 eq) and 4,4'-dimethoxytrityl chloride (1.9 g, 1.5 eq). After two hours, water (15 ml) was added and the product extracted into ethyl acetate (3 x 75 ml), dried ($MgSO_4$) and concentrated *in vacuo*. The residue was purified by silica gel chromatography (0 - 8% methanol / DCM / 2% TEA) and the product concentrated to a foam *in vacuo* (1.46 g, 66%). (F. R_f = 0.4 and 0.3). δ_H (200 MHz; $CDCl_3$), 2.15 (m, 1H, 2'H), 3.05 and 3.1 (2s, 6H, N^2NMe_2) , 3 (m, 1H, 2'H), 3.25 (m, 1H, 5'H), 3.65 (m, 1H, 5'H), 3.65 (s, 6H, DMTr-OMe), 3.8 (s, 3H, C8-OMe), 4 (br. d, 1H, 4'H), 4.27 (br, 1H 3'H), 5.24 (br, 1H, D_2O exch. 3'OH), 6.2 (t, 1H, 1'H), 6.75 (br, 4H, 3,5,3',5' DMTr-H), 7.25 (br, 9H, DMTr-H), 8.4 (s, 1H, N^2CHN), 11.33 (bs, 1H, D_2O exch. N(1)H). [Found m/z = 655.28799 $(M+H)^+$ Calc. for $C_{35}H_{39}N_6O_7$ = 655.28800]. m/z 303 $(DMTr)^+$ and 237 $(M+H\text{-}dR)^+$.

5. *5'(O)-(4,4'-Dimethoxytrityl)-N^2-[(dimethylamino)methylene]-8-methoxy-2'-deoxyguanosine-3'-(O)-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite.* To a sealed flask, flushed with argon and containing 5'(O)-(4,4'-dimethoxytrityl)-N^2-[(dimethylamino)methylene]-8-methoxy-2'-deoxyguanosine (400 mg, 610 μM) dried by evaporation of anhydrous DCM (3 x 10 ml), was added anhydrous DCM (20 ml), N,N'-diisopropylethylamine (4 eq. 0.43 ml) and 2-cyanoethyl-N,N'-diisopropylphosphoramidochloridite (1.5 eq. 0.21 ml). After stirring for thirty minutes, ethyl acetate [20 ml, dried by storing over $MgSO_4$ for 24 h. and filtering] was added and the organic phase washed with KCl (300 m*M*, 2 x 30 ml) followed by 5% sodium carbonate solution (2 x 30 ml). The organic phase was then dried ($MgSO_4$), concentrated *in vacuo* and the residue purified by silica gel chromatography, technique A (see experimental 4.2.8.), (20% ethyl acetate / DCM) and concentrated *in vacuo* to a yellow gum (455 mg, 87%). (O. $R_f = 0.25$). [^{31}P nmr. s, 148.751ppm]. δ_H (200 MHz; $CDCl_3$), 1.2 (br. m, 12H, 3'NIpMe_2), 2.3 (br. m, 1H, 2'H), 2.45 (t, 1H, 3'P-O-CHa), 2.6 (t, 1H, P-O-CHb), 3.1 (s, 6H, N^2NMe_2), 3.25 (t, 2H, 3'NC-CH_2), 3.65 (br. m, 4H, 5'H and 3'O-NCH) 3.8 (s, 6H, DMTr-OMe) 3.95 (d, 3H, C8-OMe) 4.15 (m, 1H, 4'H), 4.75 (br. m, 1H, 3'H) 6.3 (t, 1H, 1'H), 6.75 (m, 4H, trityl 2,6,2',6'H), 7.1 to 7.7 (br. m, 9H, trityl H), 8.45 (d, 1H, N^2methyleneH), 8.9 (br. s, 1H, D_2O exch. N(1)H). [Found m/z = 855.39581 $(M+H)^+$ Calc for $C_{44}H_{56}N_8O_8{}^{31}P = 855.39584$]. m/z = 637 $(M+H-C_9H_{18}O_2N_2P)^+$ and 303 $(DMTr)^+$.

6.1. *Oligonucleotide.* d[CCCCC**8MeOG**TCCCCC].

6.2 *Oligonucleotide.* CPG ----- d[C*C*C*C*C*C*8MeOG*TC*C*C*C*C*]. Synthesis was carried out as described in the experimental section but no deprotection was employed. CPG indicates that **6.2.** remained attached to the resin beads and C* and 8MeOG* represent protected bases on the oligonucleotide chain.

7. *8-Benzyloxy-2'-deoxyguanosine:-* To a stirred solution of sodium benzyloxide, prepared by addition of sodium (0.6 g, 26 mM) to purified benzyl alcohol (25 ml) and sonicating for 25 min., was added dry DMSO (15 ml) followed by 8-bromo-2'-deoxyguanosine (1.3 g, 4.3 mM) dissolved in dry DMSO (25 ml). The reaction mixture was protected from moisture and maintained at 60°C for 48 hr., with stirring, before cooling to room temperature and neutralising by the addition of Dowex (NH_4^+). After filtering the Dowex, the neutral eluent was concentrated, as far as possible, on a rotary evaporator, under oil pump vacuum, before precipitation into rapidly stirred diethyl ether (600 ml). The pale yellow precipitate was then filtered and dried *in vacuo* to yield the crude product (1.3 g, 81%). (H. R_f = 0.68). δ_H (360 MHz; d^6-DMSO) 2 (qd, 1H, 2'H), 2.85 (quin. 1H, 2'H), 3.4 (m, 2H, 5'H), 3.75 (d, 1H, 4'H), 4.25 (br, 1H, 3'H), 4.8 (t, D_2O exch. 1H, 5'OH), 5.2 (d, D_2O exch. 1H, 3'OH), 5.45 (s, 2H, C8-OCH_2), 6.1 (t, 1H, 1'H), 6.3 (s, D_2O exch. 2H, N^2H), 7.4 (br, 5H, C8-Ph), 10.6 (s, D_2O exch. 1H, N(1)H). [Found m/z = 374.14643 $(M+H)^+$ Calc for $C_{17}H_{19}N_5O_5$ = 374.14644]. UV (pH 5.6) λ_{max} = 248, 290.

9. *5',3'-O-N2-Triisobutyryl-8-oxo-7-hydro-2'-deoxyguanosine:-* 8-Benzyloxy-2'-deoxyguanosine (1 g, 2.7 mM) was dried by evaporation of anhydrous pyridine (3 x 10 ml) then suspended in anhydrous pyridine (20 ml) and cooled in an ice / water bath. Isobutyryl chloride (2.8 ml, 27mM) was then added, with stirring, and the mixture protected from moisture with a guard column. After 2 hr. the reaction was poured into a solution of 5% sodium hydrogen carbonate (50 ml) and the product extracted with ethyl acetate (3 x 50 ml), dried ($MgSO_4$) and concentrated *in vacuo* to an oil. The residue was then purified by silica gel chromatography (0 - 5% methanol / DCM) and concentrated *in vacuo* to a foam (0.94 g, 60%), compound **8.** The triisobutyryl-8-benzyloxy-2'-deoxyguanosine (0.7 g, 1.2 mM) was then dissolved in ethyl acetate / ethanol (6:1, 30 ml) and added to a beaker containing 10% Pd / C (80 mg) and nitrogen was blown over the vessel. The reaction mixture was then placed

into a hydrogenator under 50 psi. of hydrogen and shaken at room temperature for 36 hr. The mixture was then filtered twice through celite and the filtrate concentrated *in vacuo* prior to purification on silica gel chromatography (0 - 5% methanol / DCM) to yield the product (0.25 g, 42%). (K. $R_f = 0.34$). δ_H (200 MHz; $CDCl_3$), 1 (d, 6H, N^2IbMe), 2.2 (m, 1H, 2'H), 2.5 (m, 2H, 3',5'IbCH), 2.8 (sept. 1H, N^2IbCH), 3 (m, 1H, 2'H), 4.1 (m, 2H, 5'H), 4.3 (m, 1H, 4'H), 5.4 (m, 1H, 3'H), 6 (t, 1H, 1'H), 11.2 (s, D_2O exch. 1H, N(7)H), 11.5 (s, D_2O exch. 1H, N^2H), 12 (s, D_2O exch. 1H, N(1)H). [Found m/z = 494.22508 $(M+H)^+$ Calc for $C_{22}H_{32}N_5O_8$ = 494.22507]. m/z = 238 (M+H-dR)+, 168 = (M+H-dR-N^2Ib)+.

10. *5',3'-Bis-O-acetyl-8-bromo-2'-deoxyguanosine.*- Bromine solution [bromine, 4 ml, in ethanol / water (1:1 v/v, 100 ml)] was slowly added to a stirred suspension of 3',5'-*bis*-O-acetyl-2'-deoxyguanosine (5g, 14.2mM) in ethanol / water (1:1 v/v, 25 ml) until the yellow colour persisted. At this point the reaction was cooled in an ice / water bath, examined on tlc (F. Rf_{sm} 0.18, Rf_{prod} 0.3) and bromine solution added until no starting material remained. The solid was then filtered, washed with diethyl ether and recrystallised from ethanol / water (9:1 v/v, 300 ml) to give 5',3'-*bis*-O-acetyl-8-bromo-2'-deoxyguanosine (4.05 g, 66%) as small white crystals. m.pt.-darkens at 218°C and does not then melt below 350°C. [Found 39.3%C, 4.0%H, 15.8%N, Calc. 39.1%C, 3.7%H, 16.2%N]. δ_H (200 MHz; d^6DMSO). 2 and 2.1 (s, 6H, 5',3'AcMe), 2.4 (m, 1H, 2'H), 3.5 (bm, 1H, 2'H), 4.2 (m, 2H, 5'H), 4.35 (m, 1H, 4'H), 5.4 (m, 1H, 3'H), 6.2 (t, 1H, 1'H), 6.55 (s, D_2O exch. 2H, N^2H), 10.9 (s, D_2O exch. 1H, N(1)H). m/z 430 and 432 $(M+H)^+$, 447 and 449 $(M+H_2O)^+$.

11.1. *5',3'-Bis-O-acetyl-N^2-isobutyryl-8-bromo-2'-deoxyguanosine.*- 5',3'-*Bis*-O-acetyl-8-bromo-2'-deoxyguanosine (5 g, 11.6 mM), dried by evaporation of anhydrous pyridine (3 x 10 ml), was suspended in anhydrous pyridine (80 ml). The stirred suspension was protected from moisture, cooled in an ice / water bath and

isobutyryl chloride (5eq. 6 ml) added. After 15 minutes the reaction was quenched by pouring it slowly into 5% sodium hydrogen carbonate solution (200 ml) and the product was extracted into ethyl acetate (3 x 75 ml), dried ($MgSO_4$) and concentrated *in vacuo*. The residue was purified by silica gel chromatography (0 - 5% methanol / DCM) and concentrated to a foam *in vacuo* (5 g, 86%). (E. $R_f = 0.6$) δ_H (200 MHz; $CDCl_3$) 1.25 (over.d, 6H, N^2IbMe), 2.1(2s, 6H, 5',3'AcMe), 2.4 (br. m, 1H, 2'H), 2.75 (sept. 1H, N^2IbCH), 3.25 (br. m, 1H, 2'H), 4.3(m, 2H, 5'H), 4.8 (d, 1H, 4'H), 5.45 (br. d, 1H, 3'H), 6.3 (t, 1H, 1'H), 8.6 (br. s, D_2O exch. 1H, N^2H). [Found: m/z = 500.07813 (M+H)+ Calc. for $C_{18}H_{23}N_5O_7{}^{79}Br$ = 500.07812 and m/z = 502.07618 (M+H)+ Calc. for $C_{18}H_{23}N_5O_7{}^{81}Br$ = 502.07615]. m/z = 300 and 302 (M+H-dR)+.

11.2 *5',3'-Bis-O-acetyl-N^2-phenylacetyl-8-bromo-2'-deoxyguanosine*.- To a suspension of 5',3'-*bis*-O-acetyl-8-bromo-2'-deoxyguanosine (4 g, 9.3 mM), dried by evaporation of anhydrous pyridine (3 x 10 ml), in anhydrous pyridine (50 ml) was added phenylacetic anhydride (9.4 g, 4 eq.) and the mixture refluxed for 1.5 h. It was then allowed to cool to room temperature before adding saturated sodium hydrogen carbonate (50 ml) and extracting the product with dichloromethane (3 x 50 ml). The organic phase was then dried ($MgSO_4$), concentrated *in vacuo*, purified by silica gel chromatography (0 - 3% methanol / DCM) and the pure product concentrated to a foam (3.77 g, 73%). (F. $R_f = 0.62$). δ_H (360 MHz; $CDCl_3$), 2.1 (2s, 6H, 5',3'AcMe), 2.5 (m, 1H, 2'H), 3.25 (quin, 1H, 2'H), 3.8 (s, 2H, N^2PhCH_2), 4.4 (m, 2H, 5'H), 4.9 (m, 1H, 4'H), 5.5 (d, 1H, 3'H), 6.4 (t, 1H, 1'H), 7.4 (m, 5H, N^2Ph-H), 9.4 (s, D_2O exch. 1H, N^2H), 12.0 (s, D_2O exch. 1H, N(1)H). [Found: m/z = 548.07814 (M+H)+ Calc. for $C_{22}H_{23}N_5O_7{}^{79}Br$ = 548.07812 and m/z = 550.07614 (M+H)+ Calc. for $C_{22}H_{23}N_5O_7{}^{81}Br$ = 550.07615]. m/z = 348 and 350 (M+H-dR)+ and 270 (M+H-dR-Br)+.

12. *5'-O-(4,4'-Dimethoxytrityl)-N²-isobutyryl-8-bromo-2'-deoxyguanosine.*- To an ice cold solution of 5',3'-*bis*-O-acetyl-N²-isobutyryl-8-bromo-2'-deoxyguanosine (1.2 g, 2.4 mM) in pyridine / methanol / water (55:25:20, 25 ml) was added pyridine / methanol / 2 M NaOH (55:25:20, 25 ml). After five minutes the mixture was neutralised by the addition of Dowex (pyridinium form) which was then filtered and washed with pyridine / methanol (1:1). The neutral filtrate was then concentrated *in vacuo* and the paste dissolved in methanol (1 ml). The product was precipitated by addition of diethyl ether (10 ml), filtered and further washed with diethyl ether to give N²-isobutyryl-8-bromo-2'-deoxyguanosine as a white powder (0.81 g, 81%) which was dried by evaporation of anhydrous pyridine (3 x 10 ml). The solid was then suspended in anhydrous pyridine (25 ml) and triethylamine (0.55 ml, 2 eq.), 4-dimethylaminopyridine (12 mg, 0.05 eq.) and 4,4'-dimethoxytrityl chloride (0.81 g, 1.2 eq.) were added with stirring and the exclusion of moisture. After 2h. water (20 ml) was added and product extracted with ethyl acetate (3 x 35 ml), dried ($MgSO_4$) and concentrated *in-vacuo*. The resultant oil was then purified by silica gel chromatography (0 - 5% methanol / DCM / 2% TEA) and concentrated *in-vacuo* to a foam (1.1 g, 79%). (F. R_f = 0.5). δ_H (360 MHz; d⁶DMSO) 1.1 (2d, 6H, N²IbMe), 2.3 (m, 1H, 2'H), 2.75 (sept. 1H, N²IbCH), 3.1 (dd, 1H, 5'H), 3.2 (m, 1H, 2'H), 3.4 (t, 1H, 5'H) 3.7 (2s, 6H, DMTr-OMe), 4.0 (m, 1H, 4'H), 4.5 (m, 1H, 3'H), 5.2 (d, D_2O exch. 1H, 3'OH), 6.3 (t, 1H, 1'H), 6.75 (2d, 4H, 3,5,3',5' DMTr-H), 7.1 - 7.4 (m, 9H, DMTr-H), 11.3 (br. s, D_2O exch. 1H, N²H), 12.0 (br. s, D_2O exch. 1H, N(1)H). m/z = 821 and 819 $(M+H+TEA)^+$, 303 $(DMTr)^+$: m/z 716.6 and 718.6 $(M-H)^-$.

13. *5'-O-[4,4'-Dimethoxytrityl]-8-bromo-2'-deoxyguanosine:*- To a solution of 5'-O-[4,4'-dimethoxytrityl]-N²-isobutyryl-8-bromo-2'-deoxyguanosine (200 mg, 0.28 mM) in THF (5 ml) was added NaOH (0.24 g, 10 eq). This resulted in the formation of two layers and the mixture was rapidly stirred at 55°C. After 24 hr. the reaction

was complete (G. R_f = 0.28) and the mixture was then neutralised with Dowex (pyridinium form), before purifying by flash chromatography (0 - 10% methanol / DCM / 2%TEA). Concentrating *in vacuo* gave a foam (103 mg, 57%). δ_H (360 MHz; d6DMSO) 2.3 (m, 1H, 2'H), 3.1 (m, 2H, 2',5'H), 3.4 (t, 1H, 5'H) 3.7 (s, 6H, DMTr-OMe), 3.9 (bq, 1H, 4'H), 4.5 (br, 1H, 3'H), 5.2 (d, D_2O exch. 1H, 3'OH), 6.2 (t, 1H, 1'H), 6.4 (s, D_2O exch. 2H, N^2H), 6.8 (2d, 4H, 3,5,3',5'DMTr-H), 7.1-7.4 (br, 9H, DMTr-H), 10.8 (br. s, D_2O exch. 1H, N(1)H. m/z = 751 and 749 (Br present) $[M+H+TEA]^+$, m/z 303 $(DMTr)^+$. m/z = 648 $[M-H]^-$.

14. *N^2-(4,4'-Dimethoxytrityl)-8-bromo-2'-deoxyguanosine.-* 5',3'-*Bis*-O-acetyl-8-bromo-2'-deoxyguanosine (0.55 g, 1.3 mM) was dried by evaporation of anhydrous pyridine (3 x 10 ml) and dissolved in anhydrous pyridine (10 ml). The solution was protected from moisture and triethylamine (0.36 ml, 2 eq.), 4-dimethylaminopyridine (8 mg, 0.05 eq.) and 4,4'-dimethoxytrityl chloride (0.83 g, 1.2 eq) were added with stirring. After 2 h. water (10 ml) was added and the product was extracted into ethyl acetate (3 x 20 ml), dried ($MgSO_4$) and concentrated to a foam *in vacuo*. The foam was then dissolved in a solution of pyridine / methanol / water (55:25:20, 25 ml), cooled in an ice bath and pyridine / methanol / 2 M NaOH (55:25:20, 25 ml) added. After ten minutes the mixture was neutralised by the addition of Dowex, (pyridinium form), which was then filtered and washed with pyridine / methanol (1:1). The neutral eluent was concentrated *in vacuo*, the residue washed with diethyl ether, purified by silica gel chromatography (0 -15% methanol / DCM / 2% TEA) and concentrated *in vacuo* to a foam (0.68 g, 82%). δ_H (200 MHz; d6DMSO) 1.3 (bm, 1H, 2'H), 2.3 (br. m, 1H, 2'H), 3 (br. d, 1H, 5'H), 3.5 (m, 2H, 3',5'H), 3.55 (br. s, 1H, D_2O exch. N^2H), 3.75 (s, 6H, DMTr-OMe), 4. (br. s, 1H, D_2O exch. 5'OH), 5 (s, 1H, D_2O exch. 3'OH), 5.75 (t, 1H, 1'H), 6.75 - 7.3 (br, 13H, DMTr-H), 7.6 (s, 1H, D_2O exch. N(1)H). [Found: m/z = 648.14585 $(M+H)^+$ Calc. for $C_{31}H_{31}N_5O_6{}^{81}Br$ = 648.14581] m/z= 648 and 650 $(M+H)^+$, 749 and 751 $(M+H+TEA)^+$ and 303

(DMTr)+.

15.1. *5',3'-Bis-O-acetyl-N²-isobutyryl-8-(p-nitrophenylthio)-2'-deoxyguanosine.-* A solution of 5',3'-*bis*-O-acetyl-N²-isobutyryl-8-bromo-2'-deoxyguanosine (0.75 g, 1.5 mM) and p-nitrothiophenol (0.28 g, 1.2 eq.) in anhydrous pyridine (10 ml) was maintained at room temperature for 3 h before diluting with water (30 ml) and extracting with ethyl acetate (3 x 30 ml). The organic phase was then dried ($MgSO_4$), concentrated *in vacuo* and the residue purified by silica gel chromatography (0 - 3% methanol / DCM) and concentrated to a foam *in vacuo* (0.7 g, 81%). (G. R_f = 0.64), δ_H (200 MHz; $CDCl_3$) 1.25 (2d, 6H, N²IbMe), 2.15 (2s, 6H, 5',3'AcMe), 2.3 (qd, 1H, 2'H), 2.75 (sept. 1H, N²IbCH), 3.15 (quin. 1H, 2'H), 4.3 (m, 2H, 5'H), 4.9 (d, 1H, 4'H), 5.5 (br. d, 1H, 3'H), 6.5 (t, 1H, 1'H), 7.35 (d, 2H, *m*-NPT-H), 8.1 (d, 2H, *o*-NPT-H), 9.3 (s, D_2O exch. 1H, N²H), 12.2 (s, D_2O exch. 1H, N(1)H). [Found: m/z = 575.15600 $(M+H)^+$ Calc. for $C_{24}H_{27}N_6O_9S$ = 575.16601]. m/z = 375 $(M+H\text{-}dR)^+$.

15.2. *5',3'-Bis-O-acetyl-N²-phenylacetyl-8-(p-nitrophenylthio)-2'-deoxyguanosine.-* A solution of 5',3'-*bis*-O-acetyl-N²-phenylacetyl-8-bromo-2'-deoxyguanosine (0.8 g, 1.5 mM) and p-nitrothiophenol (0.28 g, 1.2 eq) in anhydrous pyridine (10 ml) was maintained at room temperature for 3 h before diluting with aqueous KCl (30 ml, 500 mM) and extracting with ethyl acetate (3 x 30 ml). The organic phase was then dried ($MgSO_4$), concentrated *in-vacuo* and the residue purified by silica gel chromatography (0 - 3% methanol / DCM) and concentrated to a foam *in vacuo* (0.94 g, 65%). (G. R_f = 0.65). δ_H (200 MHz; $CDCl_3$) 2.1 (s, 6H, 5',3'AcMe), 2.25 (qd, 1H, 2'H), 3.1 (quin. 1H, 2'H), 3.8 (s, 2H, N²PhCH₂), 4.3 (m, 2H, 5'H), 5 (d, 1H, 4'H), 5.45 (m, 1H, 3'H), 6.5 (t, 1H, 1'H), 7.4 (s, 5H, N²Ph-H), 7.4 (d, 2H, *m*-NPT-H) 8.1 (d, 2H, *o*-NPT-H), 9.45 (s, D_2O exch. 1H, N²H), 12 (s, D_2O exch. 1H, N(1)H). [Found: m/z = 623.15607 $(M+H)^+$ Calc. for $C_{28}H_{27}N_6O_9S$ = 623.15601]. m/z = 423

(M+H-dR)+.

16. *5'-O-(4,4'-Dimethoxytrityl)-N^2-isobutyryl-8-bromo-2'-deoxyguanosine-3'-O-[(2-cyanoethyl)-(N,N'-diisopropyl)]-phosphoramidite.*- After drying the 5'(O)-(4,4'-dimethoxytrityl)-N^2-isobutyryl-8-bromo-2'-deoxyguanosine (1 g, 1.4 mM) by evaporation of dry dichloromethane (3 x 10 ml) an oven dried magnetic stirrer was placed in the flask which was then sealed with a septum and flushed with argon. Dry dichloromethane (20 ml) was then added and N,N'-diisopropylethylamine (0.5 ml, 2 eq.) and 2-cyanoethyl-N,N'-diisopropylphosphoramidochloridite(0.5 ml, 2 eq.) were added to the white suspension with stirring. After 30 min. the reaction mixture was poured into a separating funnel containing dichloromethane (30 ml) and saturated sodium hydrogen carbonate (40 ml) and the reaction washed. The organic layer was then dried ($MgSO_4$), concentrated *in vacuo* and the oily residue purified by silica gel chromatography, technique A (see experimental 4.2.8.), (20% acetonitrile / DCM) and concentrated to a white foam *in vacuo* (0.9 g, 71%). [^{31}P nmr, $CDCl_3$: s, 147.5397ppm.]. (O. R_f = 0.49). m/z = 303 (DMTr)+. UV profile of enzyme digest was consistent with the synthesised 8-bromo-2'-deoxyguanosine λ_{max} = 260nm, 278(sh).

17. *Oligonucleotide.* d[TT**8BrG**TT]

18a. *Oligonucleotide.* d[CGC**8BrG**AATT**A**GCG].

18b. *Oligonucleotide.* d[CGC**8BrG**AATT**G**GCG].

18c. *Oligonucleotide.* d[CGC**8BrG**AATT**C**GCG].

18d. *Oligonucleotide.* d[CGC**8BrG**AATT**T**GCG].

18e. *Oligonucleotide.* d[CGC**A**AATT**8BrG**GCG].

19.1. *N^2-Isobutyryl-8-(p-nitrothiophenoxy)-2'-deoxyguanosine*.- To a cold solution of 5',3'-*bis*-O-acetyl-N^2-isobutyryl-8-(p-nitrophenylthio)-2'-deoxyguanosine (0.85 g, 1.5 mM) in pyridine / methanol / water (55:25:20, 25 ml) was added pyridine / methanol / 2 *M* NaOH (55:25:20, 25 ml). After five minutes the mixture was neutralised by the addition of Dowex (pyridinium form) which was then filtered and washed with pyridine / methanol (1:1). The neutral filtrate was then concentrated *in vacuo* and the paste dissolved in methanol (1 ml). The product was precipitated by addition of diethyl ether (10 ml), filtered and further washed with diethyl ether to give the title compound as a pale yellow solid (0.58 g, 80%). δ_H (200 MHz; $CDCl_3$) 1.15 (d, 6H, N^2IbMe), 2.1 (m, 1H, 2'H), 2.8 (quin. 1H, N^2IbCH), 3.1 (m, 1H, 2'H), 3.4 (m, 1H, 5'H), 3.6 (m, 1H, 5'H), 3.75 (d, 1H, 4'H), 4.4 (br. m, 1H, 3'H), 4.75 (t, D_2O exch. 1H, 5'OH), 5.25 (d, D_2O exch. 1H, 3'OH), 6.4 (t, 1H, 1'H), 7.5 (d, 2H, *m*-NPT-H), 8.25 (d, 2H, *o*-NPT-H), 11.7 (s, D_2O exch. 1H, N^2H) 12.25 (s, D_2O exch. 1H, N(1)H). [Found: m/z = 491.13490 $(M+H)^+$ Calc. for $C_{20}H_{23}N_6O_7S$ = 491.13488]. m/z = 375 $(M+H-dR)^+$.

19.2. *N^2-Phenylacetyl-8-(p-nitrophenylthio)-2'-deoxyguanosine*.- To a cold solution of 5',3'-*bis*-O-acetyl-N^2-phenylacetyl-8-(p-nitrophenylthio)-2'-deoxyguanosine (100 mg, 0.16 mM) in pyridine / methanol / water (55:25:20, 5 ml) was added pyridine / methanol / 2 M NaOH (55:25:20, 5 ml). After five minutes the mixture was neutralised by the addition of Dowex (pyridinium form) which was then filtered and washed with pyridine / methanol (1:1). The neutral filtrate was then concentrated *in vacuo* to give a solid which was washed with diethyl ether (5 ml) and filtered to give the product as a pale yellow powder (52 mg, 60%). (F. Rf = 0.6), δ_H (200 MHz; d^6DMSO) 2.1 (qd, 1H, 2'H), 3.15 (quin. 1H, 2'H), 3.5 (m, 2H, 5'H), 3.7 (m, 1H, 4'H), 3.8 (s, 2H, N^2CH_2), 4.4 (br. m, 1H, 3'H), 6.5 (t, 1H, 1'H), 7.3 (m, 5H, N^2Ph-H), 7.4 (d, 2H, *m*-NPT-H) 8.1 (d, 2H, *o*-NPT-H), 11.9 (s, D_2O exch. 1H, N^2H), 12.1 (s, D_2O exch. 1H, N(1)H). [Found: m/z = 539.13492 $(M+H)^+$ Calc. for $C_{24}H_{23}N_6O_7S$

= 539.13488]. m/z = 422 (M+H-phenylacetyl)+.

20. *5'-O-(4,4'-Dimethoxytrityl)-N²-isobutyryl-8-(p-nitrophenylthio)-2'-deoxyguanosine.*- To a solution of N²-isobutyryl-8-(p-nitrophenylthio)-2'-deoxyguanosine (0.58 g, 1.2 mM), dried by evaporation of anhydrous pyridine (3 x 10 ml), in anhydrous pyridine (20 ml) and protected from moisture was added, with stirring, triethylamine (2 eq. 0.42 ml), 4-dimethylaminopyridine (0.05 eq. 9 mg) and 4,4'-dimethoxytrityl chloride (1.2 eq. 0.61 g). After 2 h water (20 ml) was added and the product extracted with ethyl acetate (3 x 25 ml), dried ($MgSO_4$) and concentrated *in vacuo*. The residue was purified by silica gel chromatography (0 - 4% methanol / DCM / 2% TEA) and concentrated to a foam *in vacuo* (0.89 g, 95%). (F. R_f = 0.5). δ_H (200 MHz; $CDCl_3$) 0.9 (d, 6H, N²IbMe), 2 (m, 1H, N²IbCH), 2.2 (m, 1H, 2'H), 3.1-3.25 (m, 2H, 2',5'H), 3,4 (m, 1H, 5'H), 3.75 (s, 6H, DMTr-OMe), 4 (br. d, 1H, 4'H), 4.75 (br. s, D_2O exch. 1H, 3'OH), 4.9 (br, 1H, 3'H), 6.4 (t, 1H, 1'H), 6.75 (dd, 4H, 3,5,3',5' DMTr-H), 7-7.5 (br, 11H, DMTr and *m*-NPT-H), 8 (d, 2H, *o*-NPT-H). [Found: m/z = 792.25774 (M)+ Calc. for $C_{41}H_{40}N_6O_9S$ = 792.25773] m/z = 894 (M+H+TEA)+ and 303(DMTr)+.

21. *5'-O-(4,4'-Dimethoxytrityl)-N²-isobutyryl-8-(p-nitrophenylthio)-2'-deoxyguanosine-3'-O-[(2-cyanoethyl)-(N,N'-diisopropyl)]-phosphoramidite.*- After drying 5'(O)-(4,4'-dimethoxytrityl)-N²-isobutyryl-8-(p-nitrophenylthio)-2'-deoxyguanosine (0.86 g, 1.1 mM) by evaporation of dry dichloromethane (3 x 10 ml) an oven dried magnetic stirrer was placed in the flask which was then sealed with a septum and flushed with argon. Dry dichloromethane (20 ml) was then added followed by N,N'-diisopropylethylamine (0.4 ml, 2 eq.) and 2-cyanoethyl-N,N'-diisopropylphosphoramidochloridite(0.29 ml, 1.2 eq.) with stirring. After 30 min. the reaction was poured into a separating funnel containing dichloromehthane (30 ml) and saturated sodium hydrogen carbonate solution (40 ml) and the reaction

mixture washed. The organic layer was then dried ($MgSO_4$), concentrated *in-vacuo* and the oily residue purified by silica gel chromatography, technique A (see experimental 4.2.8.), (20% acetonitrile / DCM) and concentrated to a yellow foam *in vacuo* (0.82 g, 76%). (O. Rf = 0.54). [^{31}Pnmr. $CDCl_3$, d, 148.7577, 148.1670.] [Found: m/z = 993.37341 $(M+H)^+$ Calc. for $C_{50}H_{58}N_8O_{10}PS$ = 993.37339] m/z = 992 $(M)^+$, 791 $(M+H-C_9H_{18}N_2OP)^+$ and 303 $(DMTr)^+$. R_f = 0.54 (20% MeCN / DCM). UV profile of enzyme digest was consistent with that of the synthesised p-nitrophenylthio-2'-deoxyguanosine λ_{max} = 277nm.

22a. *Oligonucleotide.* d[CGC**8NPTG**AATT**G**GCG].

22a. *Oligonucleotide.* d[CGC**8NPTG**AATT**C**GCG].

22c. *Oligonucleotide.* d[CGC**8NPTG**AATT**T**GCG].

22d. *Oligonucleotide.* d[CGC**A**AATT**8NPTG**GCG].

23. *8-oxo-7-hydro-2'-deoxyguanosine:-* To a solution of 8MeodG (0.75g, 2.5mM) in pyridine (2) / methanol (1) / water (1) (30ml) was added thiophenol (4ml, 10eq.) followed by triethylamine (7ml, 20eq.) and the mixture stirred at 60°C for 24hr. Water (40ml) was then added and after cooling to room temperature the reaction mixture was washed with DCM (4 x 50ml). The aqueous layer was then evaporated to dryness yielding a yellow gum (0.93g). (A, R_{fprod} = 0.34, R_{fsm} = 0.43). δ_H (200 MHz; d^6DMSO). 1.1 (t, TEA CH_3), 1.94 (bm, 1H, 2'H), 2.5 (q, TEA CH_2) 2.90 (bm, 1H, 2'H), 3.5 (bm, 2H, 5'H), 3.75 (br. 1H, 4'H), 4.35 (br. 1H, 3'H), 4.85 (bs. D_2O ex. 1H, 5'OH), 5.20 (bs. D_2O ex. 1H, 3'OH), 6.05 (t, 1H, 1'H), 6.7 (s. D_2O ex. 2H, N^2H), 7.5 and 7.75 (br, pyridine and thiophenol), 10.6 (bs. D_2O ex. 1H, N(7)H), 11.1 (bs. D_2O ex. 1H, N(1)H). m/z = 284 $(M+H)^+$. λ_{max} (pH 7.3) 293nm and 246nm.

Appendix i.

X-ray crystallography, d[CGCQ^{6}MeGAATTTGCG].

Crystal data was collected by Dr. G. Leonard under the guidance of Dr. S. V. Phillips and Dr. W. Summers at the Department of Biophysics, Leeds University, using a Rigaku RU200 HB diffractometer, rotating anode and Nicolet-Siemens X100 area detector system.

A single crystal of dimensions 1.3 x 0.3 x 0.3 mm was mounted in a sealed glass capillary and X-ray data were collected at 277°K on a Rigaku RU200 HB diffractometer equipped with a rotating anode X-ray generator and a Nicolet-Siemens X100 area detector system. Data were processed using the XDS software package (Kabsch 1988). A total of 8543 reflections were merged to give 4481 unique reflections (R *merge* = 0.03), representing 88% of the total data to a resolution of 2.0 A. A second data set was collected on a crystal of similar dimensions using a conventional Stoe-Siemens AED2 four circle diffractometer. The resultant 3735 reflections were corrected for absorption, decomposition, Lorenz and polarisation effects and merged (R *merge* = 0.06) to yield 1953 unique reflections to 2.5 A. The two data sets were scaled and merged (R *merge* = 0.03), to give a total of 4557 unique reflections, 91% of the total data to 2.0A. The starting model for the refinement was obtained from the co-ordinates of the quasi-isomorphous d[CGCGAATTCGCG]$_2$ (Wing 1980) after idealising the geometry to allow for the slightly different unit cell (a = 25.44A, b = 40.70A, c = 65.99A, $\alpha = \beta = \gamma = 90^{o}$, space group $P2_12_12_1$). The initial stepwise rigid body refinement, using a modified version of SHELX (Sheldrick 1976) on all data from 10.0 A to 2.5 A, converged at R = 0.36. Subsequent Konnert-Hendrickson refinement (Hendrickson 1981) to 2.0 A, using NUCLSQ (16), including all reflections with $F > 2\sigma(F)$ and omitting the base

pairs G(4) - C(21) and C(9) - G(16) from the structure factor calculations converged at $\underline{R}$ = 0.44. Examination of difference Fourier and $2F_o - F_c$ maps on an Evans and Sutherland PS300 system using FRODO (Jones 1978) clearly indicated that the mispairs had Watson / Crick geometry. The bases in positions 4 - 21 and 9 - 16 were then included as dG : T base pairs to avoid biasing the position of the methyl groups of the O^{6}MedG bases and refinement was carried out using all data with $F > 3\sigma(F)$ in the range 7.0 A to 2.0 A. Positional and thermal parameters were included in the refinement which converged at $\underline{R}$ = 0.190 with the inclusion of 69 solvent molecules, all of which displayed good spherical density and hydrogen bonding geometry. The resultant $F_o - F_c$ difference maps showed substantial density in the plane of the mispairs, in a position expected for the proximal methyl group of O^{6}MedG. Further refinement with the inclusion of the O^{6}MedG base at positions 4 and 16 of the duplex converged to give $\underline{R}$ = 0.185 for 3118 reflections in the range 7.0 A to 2.0 A.

Bibliography.

Abbott, P.J. & Saffhill, R., (1979) *Biochem. Biophys. Acta.* **562.** 51-56.

Agrawal, S., (1992), *Trends in Biotechnology.* **10.** 152-158.

Aida, M. and Nishimura, S., (1987) S., *Mutation Res.* **192.** 83-89.

Alberts, B. et al. (1989), in "*Molecular biology of the Cell.*" 2nd Ed. Garland publishing inc. N.Y. p13.

Altman, S., (1990), *Angew. Chem. Int. Ed. (Engl.)* **29.** 759-758.

Altman, S., in "*Transfer RNA.*" MIT press, Chapter 7, 168-188. Novel ribosides.

Ames, B.N., (1979), *Science.* **204,** 587-593.

Arnott, S., Chandrasekaran, R., Huilins, D.W.L., Smith, P.J.C. and Watts, L., (1974), *J. Mol. Biol.* **88.** 523-533.

Aruoma, O.I. and Halliwell, B., (1991), *Chemistry in Britain* **27.** 149-152.

Atkins, P.W., (1986), *Physical Chemistry.* 3rd Ed. Oxford University Press. 585-586.

Beaucage, S. L. and Iyer, P., (1992), *Tetrahedron* **48.** 2223-2311.

Beaucage, S.L. and Caruthers, M., (1981) *Tetrahedron Lett.* **22**(20). 1859-1862.

Bhanot, O.S. & Ray, A., (1986), *Proc. Natl. Acad. Sci. (U.S.A.).* **83,** 7348-7352.

Blackburn, G. A. and Kellard, B., (1986), *Chemistry and Industry* 607-613.

Bodepudi, V., Shibutani, S. and Johnson, F., (1992), *Chem. Res. Toxicol.* **5.** 608-617.

Borowy-Borowski, H. & Chambers, R.W., (1987), *Biochemistry* **26,** 2465-2471.

Brown, T., Booth. E. D., Leonard, G.A. and Chambers. J., (1989), *J. Mol. Biol.* **207.** 455-457.

Brown, T., Hunter, W.N., Kneale, G. and Kennard. O., (1986), *Proc. Natl. Acad. Sci. (U.S.A.)* **83.** 2402-2406.

Bugg, C.E. and Thewalt,U., (1969), *Biochem. Biophys. Res. Commun.* **37.** 623-629.

Bugg, C.E., Thewalt,U. Marsh, R.E., (1968), *Biochem. Biophys. Res. Commun.* **33.** 436-440.

Catty, D., (ed.) (1988), *Antibodies I.: a practical approach.* IRL press limited. Oxford.

Cech, T.R. and Bass, B.L., (1986), *Ann. Rev. Biochem.* **55.** 599-620.

Cech, T.R., (1990), *Angew. Chem. Int. Ed. (Engl.)* **29.** 759-768.

Chambers, R.W., Sledziewska-Gojska, E., Hirani-Hojatti, S. & Borowy-Borowski, H., (1985), *Proc. Natl. Acad. Sci. (U.S.A.)* **82,** 7173-7177.

Chargaff, E., (1950), *Experientia,* **6.** 201.

Cheng, S.H. etal. (1990), *Cell,* **63.** 827-834.

Cheul-Hee, K., Xiaohue, Z., Ratliff, R., Moyzis, R. and Rich, A., (1992), *Nature* **356.** 126-131.

Cho, B. P., Kadlubar, F.F. Culp, S.J., and Evans, F. E., (1990), *Chem. Res. Toxicol.* **2.** 416-422.

Chung, H-J. et al. (1991), *Biochem. Biophys. Res. Commun.* **178.**(3) 1472-1478.

Culp, S.J., Cho, B. P., Kadlubar, F.F. and Evans, F. E., (1989), *Chem. Res. Toxicol.* **3.** 445-452.

Demple, B., Jacobsson, A., Olsson, M., Robins, P. & Lindahl, T., (1982), *J. Biol. Chem.* **257.**(22) 13776-13780.

Demple, B., Jacobsson, A., Olsson, M., Robins, P. & Lindahl, T., (1982), *J.Biol.Chem.* **257**(22). 13776-13780.

Dickerson, R.E., (1983), *Scientific American* **249.**(6) 94-111.

Dizdaroglu, M., Sonntac, C von. and Schulte-Frohlinde, D., (1975), *J. Am. Chem. Soc.* **97.**(8) 2277-2278.

Doi, J. T., Luehr, G. W. and Musker, W. K. (1985), *J. Org. Chem.* **50.** 5716 - 5719.

Drew, H. R. and Dickerson, R. E., (1981), *J. Mol. Biol.* **151.** 535-556.

Elion, G.R., (1989), *Angew. Chem. Int. Ed. (Engl.)* **28.** 870-878.

Englisch, U. and Gauss, D. H., (1991), *Angew. Chem. Int. Ed. (Engl.)* **30.** 613-629.

Fersht, A.R., (1982), *J. Mol. Biol.* **156.** 37-51.

Fliess, A., Wolfes, H., Seela, F. and Pignoud, A., (1988), *Nucleic Acids Res.* **16.**(11) 781

Fowler, R.G., Degnen, E. and Cox, E.C., (1974), *Mol. Gen. Genet.* **133.** 179-191.

Frieberg, E. C., (1985), in "*DNA Repair*" W. H. Freeman and company. U.S.A. Chapter 3.

Gaffney, B. L., and Jones, R.A., (1982), *Tetrahedron Letts.* **23**(22). 2257 - 2260.

Gaffney, B. L., Marky, L.A. and Jones, R.A., (1984), *Biochemistry* **23.** 5686-5691.

Gaffney, B. L. and Jones, R. A., (1989), *Biochemistry* **28.** 5881 - 5889.

Gait, M. J., (1984), in "*Oligonucleotide Synthesis: A practical approach.*" Chapter 2. p23. IRL Press Limited. Oxford.

Gao, X. and Patel, D.J., (1988), *J. Am. Chem. Soc.* **110.** 5178-5182.

Garcia-Belido, A., Lawrence, P. A. and Morata, G., (1979), *Scientific American* **241**(1). 90 - 109.

Ginell, L., Narendra, N., Jones, R.A. & Berman, H,M., (1990), *Biochemistry* **29**(46). 10461-10465.

Glick, G., (1991), *J. Org. Chem.* **56.** 6746-6747.

Graves, R.J., Li, B.F.L. and Swann, P.F., (1989), *Carcinogenesis* **10.**(4) 661-666.

Gregory, R.J., (1991), *Mol. Cell. Biol.* **11.** 3886-3893.

Guy, A., Duplaa, A-M., Harel, P. and Teoule, R., (1988), *Helv. Chim. Acta* **71.** 1566-1571.

Hanlon, S., (1966), *Biochem. Biophys. Res. Commun.* **23.** 861-867.

Hendrickson, W.A. & Konnert, J.H., (1981), *in "Biomolecular Structure, Conformation, Function and Evolution", Ed: Srinivasan, R., Pergamon Press, Oxford.* **Vol1,** 43-57.

Herskovits, (1963), *Biochemistry* **2.** 335-340.

Holmes, R. E. and Robins, R. K., (1964), *J. Am. Chem. Soc.* **86**(1). 1242-1246.

Hong, J-I., Feng, Q., Rotello, V. and Reebeck Jr. J., (1992), *Science* **255.** 848-850.

Howard, F.B., Chen, C-Q., Ross, P.D. and Miles, H.T., (1991), *Biochemistry* **30.** 779-782.

Hunter, W.N., Brown. T., Anand, N.N. & Kennard. O., (1986), *Nature.* **320.** 552-555.

Hunter, W. N., Brown, T. and Kennard, O., (1987), *Nucleic Acids Res.* **15.** 6589 - 6606.

Jones, T.A., (1978), *J. Applied Cryst.* **11,** 268-272.

Joyce, C., (1990), *New Scientist.* **127.** 24-25.

Joyce, C .M. and Steitz, T. A., (1987), *Trends in Biol. Sci.* 288 - 292.

Joyce, G.F., (1989), *Nature* **338.** 217-224.

Kabsch, W., (1988), *J. Applied Cryst.* **21,** 916-924.

Kang, C. etal. (1992), *Nature* **356.** 126-131.

Karran, P. and Marrinus, M.G., (1982), *Nature* **296.** 868-867.

Kasai, H. et al. (1987), *Bull. Chem. Soc. Japan.* **60.** 3799-3800.

Kasai, H. & Nishimura, S., (1984a), *Nucleic Acids Res.* **12.**(4) 2137-2145.

Kasai, H. & Nishimura, S., (1984b), *Gann* **75.** 565-566.

Kasai, H. & Nishimura, S., (1984c), *Gann.* **75.** 841-844.

Kennard, O. and Hunter, W.N., (1991), *Angew. Chem. Int. Ed. (Engl.)* **30.** 1254-1277.

Khoda, K. & Tomita, Ken-ichi., (1988), *Nucleic Acids Res.* **16.**(19) 9307-9321.

Knowles, F.G.W., (1945), *Man and other living things,* G.G.Harrap and Co. Ltd.

Kopka, M. L., Fratini, A. V., Drew, H. R. and Dickerson, R. E., (1983), *J. Mol. Biol.* **163.** 129-146.

Kouchakdjian, M., Bodepudi, V., Shibutani, S., Eisenberg, M., Grollman, A. P. and Patel, D. J., (1991), *Biochemistry* **30.** 1403-1412.

Kuchino, Y., Mori, F., Kasai, H., Inoue, H., Iwai, S., Miura, K., Ohtsuka, E. and Nishimura, S., (1987), *Nature* **327.** 77-79.

Kuzmich, S., Marky, L.A. & Jones, R.A., (1983), *Nucleic Acids Res.* **11,** 3393-3404.

Lau, M.M.H. and Neufeld, E.F., (1989), *J. Mol. Biol.* **11.** 3886-3893.

Leonard, G.A., Brown, T. and Hunter, W.N., (1992a), *Eur. J. Biochemistry.* **204.** 69-74.

Leonard, G. A., (1992b), *Biochemistry* **31.** 8415-8420.

Letsinger, R. L. and Lunsford, W. B., (1976), *J. Am. Chem. Soc.* **98.** 3655-3661.

Li, B. F. L. and Swann, P. F., (1989), *Biochemistry* **28.** 5779-5786.

Lippard, S. J. and Hoeschele, J. D., (1979), *Proc. Natl. Acad. Sci. (U.S.A.).* **76**(12). 6091 - 6095.

Loeb, L.A. and Kunkel, T.A., (1981), *Ann. Rev. Biochem.* **52.** 429.

Loechler, E.L, Green, C.L. & Essigman, J.M., (1984), *Proc. Natl. Acad. Sci. (U.S.A.).* **81,** 6271-6275.

Long, R. A., Robins, R. K. and Townsend, L. B., (1967), *J. Org. Chem.* **32.** 2751-2756.

Malins, D.C. and Haimanot, R., (1990), *Biochem. Biophys. Res. Commun.* **173.**(2) 614-619.

Marky, L.A. and Breslauer, K.J. (1987) Biopolymers 26. 1601 - 1620.

Matteucci, M.D. and Caruthers, M., (1981), *J. Am. Chem. Soc.* **103.** 3185-3191.

Matteucci, M., *Tetrahedron Lett.* (1990), **31.** 2385-2388.

McBride, T. J., Preston, B.D. and Loeb, L.A., (1991), *Biochemistry* **30.** 207-213.

McCloskey, J.A. and Nishimura, S., (1977), *Acc. Chem. Res.* **10.** 403-410.

Meffert, R., Rathgelser, G., Schafer, H-J. and Dose, K., (1990), *Nucleic Acids Res.* **18.**(22) 6633-6636.

Miller, S.L., (1953), *Science* **117.** 528.

Miller, E.C. & Miller, J.A., (1981), *Cancer.* **47,** 2327-2345.

Mitra, G., Pauly, G.T., Kumar, R., Pei, G.K., Hughes, S.H., Moschel, R.C. & Barbacid, M., (1989), *Proc. Natl. Acad. Sci. (U.S.A.).* **86,** 8650-8654.

Montagnier, L., (1988), *Scientific American* **259.**(4) 25-32.

Moser, H. E. and Dervan, P. B., (1987), *Science* **238.** 645-650.

Nagata, C. & Aida, M., (1988), *J. Molec. Struct. (Theochem).* **179.** 451-466.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. and Klug, A., (1987), *Nature* **330.** 221-226.

Nikiforov, T. and Connolly, B. A., (1991), *Tetrahedron Lett.* **32.** 3851-3854.

Oda, Y., et al., (1991), *Nucleic Acids Res.* **19**(7). 1407-1412.

Oliver, S.G. etal. (1992), *Nature* **357.** 38-46.

Olsson, M. & Lindahl, T., (1980), *J.Biol.Chem.* **255.**(22) 10569-10571.

Ornstein, R.L., Rein, R., Breen, D.L. and MacElroy, R.D., (1978), *Biopolymers* **17.** 2341-2360.

Ott, J. and Eckstein, F., (1987), *Biochemistry* **26.** 8237-8241.

Parsatharathy, R. & Fridey, S.M., (1986), *Carcinogenesis* **7.**(2) 221-227.

Patel, D.J., Shapiro, L., Kozlowski, S.A., Gaffney, B.L. & Jones, R.A., (1986a), *Biochemistry.* **25.** 1027-1036.

Patel, D.J., Shapiro, L., Kozlowski, S.A., Gaffney, B.L. & Jones, R.A., (1986b), *Biochemistry.* **25.** 1036-1042.

Pedersen, L.G., Darden, T.A., Deerfield II, D.W., Anderson, M.W. & Hoel, D.G., (1988), *Carcinogenesis.* **9,** 1553-1562.

Piccirilli, J.A., Krauch, T., Moroney, S.E.and Benner, S.A., (1990), *Nature* **343.** 33-37.

Porschke, D., (1977), *Mol. Biol. Biochem. Biophys.* **24.** 191-218.

Preston, B.D., Poiesz, P.J. and Loeb. L.A., (1988), *Science* **242.** 1168.

Rabinovich, D., Haran, T., Eisenstein, M. and Shakhed, Z., (1988), *J. Mol. Biol.* **200.** 151-161.

Reddy, E.P., Reynolds, R.K., Santos, E. & Barbacid, M., (1982), *Nature.* **300,** 149-152.

Riordan, L.R. and Martin, J.C., (1991), *Nature* **350.** 442-443.

Roelen, H. C. P. F., Saris, C. P., Brugghe, H. F., van den Elst, H., Westra, J. G., van der Marel, G. A. and van Boom, J. H., (1991), *Nucleic Acids Res.* **19**(16). 4361 - 4369.

Rose, R.C., (1990), *Biochem. Biophys. Res. Commun.* **169.**(2) 430-436.

Saenger, W., (1984), in "*Principles of Nucleic Acid Structure.*" Springer-Verlag.

Saffhill, R., Margison, G.P., & O'Connor, P., (1985), *Biochim. Biophys. Acta.* **823,** 111-145.

Sanger, F., Nicklen, S. and Coulson, A.R., (1977), *Proc. Natl. Acad. Sci. (U.S.A.).* **74.**(12) 5463-5467.

Scicchitano, D., Jones, R.A., Kuzmich, S., Gaffney, B.L., Lasko, D.D., Essigman, J.M. & Pegg, A.E., (1986), *Carcinogenesis* **7.**(8) 1383-1386.

Searle, C.E., (1986), *Chemistry in Britain.* 211-220.

Sheldrick, G.M., (1976), *Shelx 76 system of Computing Programs, Univ. Cambridge, UK.*

Shibutani, S., Takeshita, M. and Grollman, A. P., (1991), *Nature* **349.** 431-434.

Shigenaga, M.K., Gimeno, G.J. & Ames, B.N., (1989), *Proc. Natl. Acad. Sci. (U.S.A.).* **86.** 9697-9701.

Singer, B., (1982a), *Basic life science.* **20.** 1-42.

Singer, B. & Kusmierk, J.T., (1982b),*Ann. Rev. Biochem.* **52.** 655-693.

Smith A., (1975), *The Human pedigree* Allan and Unwin Ltd. Chapter 6. p123-125.

Snow, E.T., Foote, R.S. & Mitra, S., (1984), *J. Biol. Chem.* **259.**(13) 8095-8100.

Sood, A., Shaw, B.R. and Spielvogel, B.F., (1990), *J. Am. Chem. Soc.* **112.** 9000-9001.

Stein, C.A. etal. (1991), *Biochemistry* **30.** 2439-2444.

Stubbe, J., (1988), *Biochemistry* **27.**(11) 3893-3900.

Sukumar, S., Notario, V., Martin-Zanca, D. and Barbacid, M., (1983), *Nature* **306.** 658.

Suttcliffe, E. Y. and Robins, R. K., (1963), *J. Org. Chem.* **28.** 1662-1666.

Tai-Shun Lin, Jia-Chong Chen, Ishiguro, K. and Sartorelli, A. C., (1985), *J. Med. Chem.* **28.** 1194-1198.

Taylor, J. W., Ott, J. and Eckstein, F., (1985), *Nucleic Acids Res.* **13.** 8765.

Taylor, J.D. and Halford, S.R., (1989), *Biochemistry* **28.** 6198.

Taylor, J.D., Badcoe, I.G., Clarke, A.R. and Halford, S.R., (1991), *Biochemistry* **30.** 8743.

Toorchen, D. & Topal, M.D., (1983), *Carcinogenesis* **4.**(12) 1591-1597.

Uesugi, S. & Ikehara, M., (1977), *J. Am. Chem. Soc.* **99.**(10). 3250-3253.

Uhlmann, E. and Peyman, A., (1990), *Chemical Revs.* **90.**(4) 543-584.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, C., (1991), *Science* **243.** 1503-1507.

Wallwork, S.C., (1961), *J. Am. Chem. Soc.* 494.

Watson, J.D. and Crick, F.H.C., (1953), *Nature* **171.** 737.

Wells, R.D. etal. (1988), *FASEB. J.* **2.** 2939-2949.

Will, D. W. and Brown, T., (1992), *Tetrahedron Lett.* **33.**(19) 2729-2732.

Williams, L.D. & Shaw, B.R., (1987), *Proc. Natl. Acad. Sci. (U.S.A.).* **84,** 1779-1783.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E., (1980), *Nature* **287.** 755 - 758.

Yamagata, Y., Kohda, K. & Tomita, K., (1988), *Nucleic Acids Res.* **16,** 9307-9321.

Yarchoan, R., Mitsuya, H. and Broder, S., (1988), *Scientific American* **259.**(4) 88-97.

Ying Li, Zon, G. and Wilson, W.D., (1991), *Proc. Natl. Acad. Sci. (U.S.A.).* **88.** 26 - 30.

Zarbl, H., Sukumar, S., Arthur, A.V., Martin-Zanka, D. & Barbacid, M., (1985), *Nature.* **315,** 382-385.

Publication.

High-resolution structure of a mutagenic lesion in DNA

(x-ray diffraction/ultraviolet melting/guanine methylation)

G. A. LEONARD*, J. THOMSON*, W. P. WATSON†, AND T. BROWN*‡

*Department of Chemistry, Edinburgh University, Kings Buildings, EH9 3JJ Scotland, United Kingdom; and †Shell Research Ltd, Sittingbourne Research Centre, Kent ME9 8AG, United Kingdom

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9573–9576, December 1990
Biochemistry

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9573–9576, December 1990
Biochemistry

High-resolution structure of a mutagenic lesion in DNA

(x-ray diffraction/ultraviolet melting/guanine methylation)

G. A. LEONARD*, J. THOMSON*, W. P. WATSON†, AND T. BROWN*‡

*Department of Chemistry, Edinburgh University, Kings Buildings, EH9 3JJ Scotland, United Kingdom; and †Shell Research Ltd, Sittingbourne Research Centre, Kent ME9 8AG, United Kingdom

Communicated by John Cornforth, September 4, 1990 (received for review June 26, 1990)

ABSTRACT **The self-complementary dodecanucleotide d[CGC(m^{6}G)AATTTGCG]$_2$ (where m^{6}G is O^6-methylguanine), which contains two m^{6}G·T base pairs, has been analyzed by x-ray diffraction methods and the structure has been refined to a residual error of $R = 0.185$ at 2.0-Å resolution. The m^{6}G·T mispair closely resembles a Watson–Crick base pair and there are very few structural differences between the m^{6}G·T duplex and the native analogue. The similarity between the m^{6}G·T base pair and a normal G·C base pair explains the failure of mismatch repair enzymes to recognize and remove this mutagenic lesion. A series of ultraviolet melting studies over a wide pH range on a related dodecamer indicate that the m^{6}G·C mispair can exist in two conformations; one is a wobble pair and the other is a protonated Watson–Crick pair. The former, which predominates at physiological pH, will be removed by normal proofreading and repair enzymes, whereas the latter is likely to escape detection. Hence, the occasional occurrence of the protonated m^{6}G·C base pair may explain why the presence of m^{6}G in genomic DNA does not always give rise to a mutation.**

The initial stages of chemical carcinogenesis frequently involve the interaction of genotoxic agents with DNA to produce covalent modifications in the form of DNA adducts (1–3). An important example of this is the alkylation of the O^6 position of guanine residues in DNA resulting from exposure to methylating agents such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (4) and methyl methanesulfonate and *N*-methyl-*N*-nitrosourea (5). The presence of O^6-methylguanine (m^{6}G) constitutes a mutagenic lesion that is known to specifically induce G·C to A·T transition mutations (6) and it has been established that protooncogenes can be converted to oncogenes by such a process (7). Hence, the formation of the m^{6}G·T base pair during replication can give rise to a carcinogenic lesion (8, 9). In recent years, the biochemical processes involved in chemically induced carcinogenesis have been studied in considerable depth. However, to understand further the mechanisms of mutagenesis, it is necessary to analyze precisely the molecular details of the lesions produced when genotoxic agents interact with DNA. With this overall objective in mind, we have determined the structure of such a lesion, the m^{6}G·T base pair in a B-DNA duplex.§

MATERIALS AND METHODS

All oligonucleotides were synthesized by the solid-phase method on an ABI model 380B DNA synthesizer using cyanoethyl phosphoramidite monomers. For those containing m^{6}G, the following protocol was observed: The 5'-dimethoxytrityl-N^2-isobutyryl-O^6-methyldeoxyguanosine 3'-cyanoethyl phosphoramidite monomer was utilized to introduce O^6-methyldeoxyguanosine and the fully assembled oligonucleotide was cleaved from the solid support and deprotected in a 5% solution of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in anhydrous methanol for 2 weeks at ambient temperature in an atmosphere of nitrogen (10). At no time was the oligonucleotide exposed to ammonia, as this can lead to the slow conversion of m^{6}G to 2,6-diaminopurine (11). DBU was removed with Dowex-50 cation-exchange resin (Na^+ form) and purification was accomplished by reversed-phase HPLC (octyl), eluting with a linear gradient of acetonitrile in 0.1 M ammonium acetate (0%–20%, 30 min) to give a major and a minor product in approximate ratios of 9:1. These were easily separated due to the much greater retention time of the latter. A sample of each was digested to the free nucleosides with snake venom phosphodiesterase and alkaline phosphatase and the base composition was analyzed by reversed-phase HPLC (10). In all cases, this confirmed that the oligonucleotide containing O^6-methyldeoxyguanosine was the major product. The minor product was found to contain N^2-isobutyryl-O^6-methylguanine, and it could be converted to the desired product by further treatment with DBU in methanol. Before carrying out pH-dependent ultraviolet melting studies, the stability of m^{6}G at low pH was determined in the following way: The oligonucleotide d[CGC(m^{6}G)AATTCGCT] was dissolved in 0.1M sodium phosphate (pH 5.0) and, after 2 weeks, reversed-phase HPLC analysis showed that there was no significant degradation. Mixed injections with the native sequence d(CGCGAATTCGCG), which elutes much earlier on reversed-phase HPLC, further confirmed that the m^{6}G-containing oligonucleotide had not undergone demethylation.

The self-complementary dodecamer d(CGCm^{6}GAATTTGCG)$_2$ crystallized isomorphously with the native dodecamer (12). Crystals were grown at 277 K and pH 6.3 from 25-μl drops containing 1.0 mM oligonucleotide, 33 mM $MgCl_2$, 16.7% (vol/vol) hexylene glycol, 1.25 mM spermine, and 5 mM sodium cacodylate. A single crystal of dimensions 1.3 × 0.3 × 0.3 mm was mounted in a sealed glass capillary and x-ray data were collected at 277 K on a Rigaku RU200 HB diffractometer equipped with a rotating anode x-ray generator and a Nicolet–Siemens X100 area detector system. Data were processed using the XDS software package (13). A total of 8543 reflections were merged to give 4481 unique reflections (R merge = 0.03), representing 88% of the total data to a resolution of 2.0 Å. A second data set was collected on a crystal of similar dimensions using a conventional Stoe–Siemens AED2 four circle diffractometer. The resultant 3735 reflections were corrected for absorption, decomposition, Lorentz and polarization effects and merged (R merge = 0.06) to yield 1953 unique reflections to 2.5 Å. The two data sets were scaled and merged (R merge = 0.03) to give a total of 4557 unique reflections, 91% of the total data to 2.0 Å. The starting model for the refinement was

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: m^{6}G, O^6-methylguanine.

‡To whom reprint requests should be addressed.

§The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1D27, R1D27SF).

obtained from the coordinates of the quasiisomorphous d(CGCGAATTCGCG)$_2$ (12) after idealizing the geometry to allow for the slightly different unit cell (a = 25.44 Å; b = 40.70 Å; c = 65.99 Å; $\alpha = \beta = \gamma = 90°$; space group $P2_12_12_1$). The initial stepwise rigid body refinement, using a modified version of SHELX (14) on all data from 10.0 Å to 2.5 Å, converged at R = 0.36. Subsequent Konnert–Hendrickson refinement (15) to 2.0 Å, using NUCLSQ (16), including all reflections with $F > 2\sigma(F)$ and omitting the base pairs G(4)–C(21) and C(9)–G(16) from the structure factor calculations converged at R = 0.44. Examination of difference Fourier and $2F_o - F_c$ maps on an Evans and Sutherland PS300 system using FRODO (17) clearly indicated that the mispairs had Watson–Crick geometry. The bases in positions 4–21 and 9–16 were then included as G·T base pairs to avoid biasing the position of the methyl groups of the m^{6}G bases and refinement was carried out using all data with $F > 3\sigma(F)$ in the range 7.0 Å to 2.0 Å. Positional and thermal parameters were included in the refinement that converged at R = 0.190 with the inclusion of 69 solvent molecules, all of which displayed good spherical density and hydrogen-bonding geometry. The resultant $F_o - F_c$ difference maps showed substantial density in the plane of the mispairs, in a position expected for the proximal methyl group of m^{6}G. Further refinement with the inclusion of the m^{6}G base at positions 4 and 16 of the duplex converged to give R = 0.185 for 3118 reflections in the range 7.0 Å to 2.0 Å.

RESULTS AND DISCUSSION

The x-ray structure analysis of the self-complementary dodecanucleotide d[CGC(m^{6}G)AATTTGCG]$_2$ shows that in overall shape the m^{6}G·T mispair is remarkably similar to a Watson–Crick base pair. The bases are directly opposite each other, the glycosyl linkages are related by a pseudodyad, and in the minor groove the base pair is indistinguishable from a G·C base pair. Overall, there are very few structural differences between the m^{6}G·T duplex and the well-studied native G·C duplex, d(CGCGAATTCGCG)$_2$ (12). The helical parameters, torsion angles, and hydration pattern are almost identical and the presence of the mutagenic base pair does not distort the sugar-phosphate backbone. Despite these considerable similarities, ultraviolet melting studies indicate that the m^{6}G·T duplex is much less stable than the native G·C duplex ($\Delta\Delta G°$ = 46.8 kJ/mol) (Table 1). Hence, the m^{6}G·T base pair has a powerful destabilizing effect in B-DNA.

The two m^{6}G·T base pairs in the dodecamer duplex are essentially identical and the $2F_o - F_c$ map of the m^{6}G(9)·T (16) pair is shown in Fig. 1*a*. There are three close contacts between the m^{6}G and T bases; one of 2.9 Å from purine-O^6 to pyrimidine-O^4, a second of 2.9 Å from purine-N^1 to pyrimidine-N^3, and a third of 2.8 Å between purine-N^2 and pyrimidine-O^2. At this resolution, we cannot observe hydrogen atoms and the most likely base pair is shown in Fig. 2, with each base as its most common tautomer. There is unlikely to be a hydrogen bond between purine-O^6 and pyrimidine-O^4, as neither has an attached hydrogen and neither functional group is sufficiently basic to be protonated at pH 6.3. In addition, there is no evidence in the pH-dependent ultraviolet melting profile of the duplex (Fig. 3*a*) for a protonated base pair. Hence, formation of the double strand from the fully hydrated single strands will lead to the loss of a hydrogen bond between each of these oxygen atoms and neighboring solvent molecules. As these are not replaced by interbase hydrogen bonds, the effect will be to destabilize the duplex. The interaction between purine-O^6 and pyrimidine-O^4 will be destabilizing for the additional reason that the two electronegative oxygen atoms are forced together by the two adjacent strong hydrogen bonds in the base pair (20), one from pyrimidine-N^3 to purine-N^1 and a second from purine-N^2 to pyrimidine-O^2. It is possible to postulate other forms of the m^{6}G·T base pair by invoking minor tautomers but there is no direct experimental evidence for their existence. The base pair in Fig. 2 has two hydrogen bonds that are adjacent to each other and the resultant cooperativity allows the formation of a stable base pair. The m^{6}G·T base pairing found in the present x-ray structure is identical to that postulated from molecular orbital and molecular mechanical calculations and from an NMR study of a mixture of ribonucleosides of m^{6}G and thymine in chloroform solution (21). However, it differs in important details from that proposed in an oligonucleotide NMR study, although interpretation of the spectra was limited by the lack of direct information on the relative orientation of the two bases (22).

The most striking differences between the m^{6}G·T mispair and a G·C base pair are in the major groove, due to the presence of the methyl group attached to the purine-O^6 atom, located proximal to the N^7 atom in the plane of the purine ring (Fig. 1*b*). This will give rise to steric repulsion between the guanine base and the attached methyl group. Although the distal conformation is preferred in the free nucleoside (23, 24), it is likely to be very unstable in the m^{6}G·T base pair (25) as it will prevent the formation of interbase hydrogen bonds. In the proximal orientation, the methyl group presents a steric barrier to any regulatory or repair enzyme that might otherwise interact with the guanine-O^6 or -N^7 atom. The appearance of the m^{6}G·T base pair in the major groove is different from that of a Watson–Crick A·T or G·C base pair in two additional details: (*i*) there are two methyl groups, one attached to each base, and (*ii*) there is no heteroatom with a capacity to donate hydrogen bonds.

When a guanine base in genomic DNA is converted to m^{6}G by a chemical mutagen, the modified base normally codes for thymine instead of cytosine (26), resulting in an *in vivo* mutation frequency of between 15% (27) and 75% (28). The *in vitro* misinsertion frequency is >95% (29) and the only efficient form of repair involves demethylation of the O^6 atom of guanine by the enzyme m^{6}G methyltransferase to regenerate guanine (30, 31). Thus, the m^{6}G·T base pair is recognized as being more similar to a Watson–Crick base pair than is the m^{6}G·C base pair. The reasons for this are unlikely to be thermodynamic in origin (18, 19), as ultraviolet melting

Table 1. Thermodynamic parameters for d(CGCXAATTYGCG)$_2$ duplexes at pH 7.0 in 1.0 M NaCl/10 mM sodium phosphate/1 mM EDTA

Base pair	$\Delta H°$, kJ·mol^{-1}	$\Delta S°$, J·mol^{-1}·K^{-1}	$\Delta G°$, kJ·mol^{-1}	t_m, K
G·C	−430.1	−1164	−83.2	344.6
A·T	−399.0	−1085	−75.7	341.6
G·T	−331.7	−937.6	−52.3	324.4
m^{6}G·C	−213.7	−592.3	−37.2	317.1
m^{6}G·T	−208.4	−577.3	−36.4	315.2

Thermodynamic parameters were determined from the concentration dependence of ultraviolet melting by standard methods (18, 19). Each point on the curve was measured in triplicate. t_m, melting temperature at 40 μM oligonucleotide; X, any nucleotide; Y, any nucleotide except that represented by X.

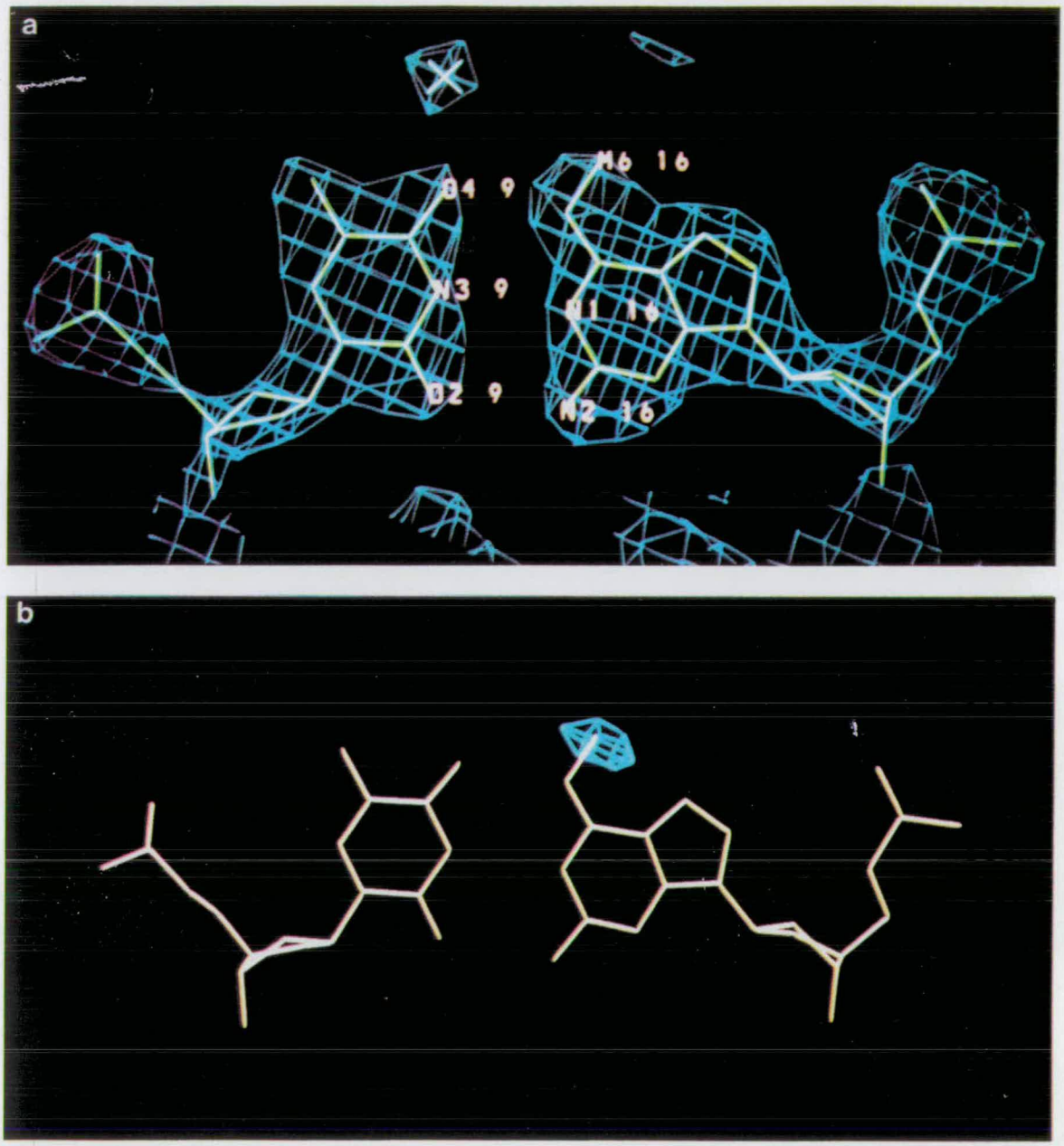


FIG. 1. (*a*) $2F_o - F_c$ map of the m^{6}G(9)·T(16) base pair. Some key atoms are labeled and a solvent molecule is highlighted in the major groove, within hydrogen bonding distance of thymine-O^4. (*b*) $F_o - F_c$ difference density map of the m^{6}G(9)·T(16) base pair with the guanine methyl group removed from the structure factor calculations. The methyl group is located in the proximal conformation in the plane of the purine ring.

studies show that the duplex containing the m^{6}G·T base pair is slightly less stable than the corresponding m^{6}G·C duplex at neutral pH ($\Delta\Delta G^\circ$ = 0.8 kJ/mol) (Table 1). Moreover, the m^{6}G·T duplex is less stable than the duplex containing the G·T wobble base pair mismatch, which is rarely incorporated during replication due to efficient proofreading ($\Delta\Delta G^\circ$ = 15.9 kJ/mol).

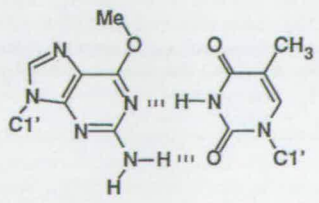


FIG. 2. m^{6}G·T base pair with each base as its major tautomer.

The incorporation of the m^{6}G·T base pair in preference to the m^{6}G·C base pair in genomic DNA can be rationalized in structural terms. The similarity in shape between the m^{6}G·T mispair and a Watson–Crick base pair, particularly in the minor groove is striking, whereas the m^{6}G·C base pair has been postulated on the basis of theoretical and NMR studies to be a reverse wobble base pair as in Fig. 4*a* (32, 25). This would, by analogy with mismatch base pairs, be removed by proofreading. Thus the enzymes in the cell nucleus responsible for DNA synthesis and repair discriminate in favor of the mutagenic lesion. We have studied the stability profile of the DNA duplex containing the m^{6}G·C base pair over a wide pH range (Fig. 3*b*) and we have observed that the melting temperature falls slightly from pH 8.5 to pH 6.5, as would be expected for the unprotonated wobble base pair 4a. At lower pH however, there is a clear indication of duplex stabilization, probably due to protonation of cytosine-N^3 of the

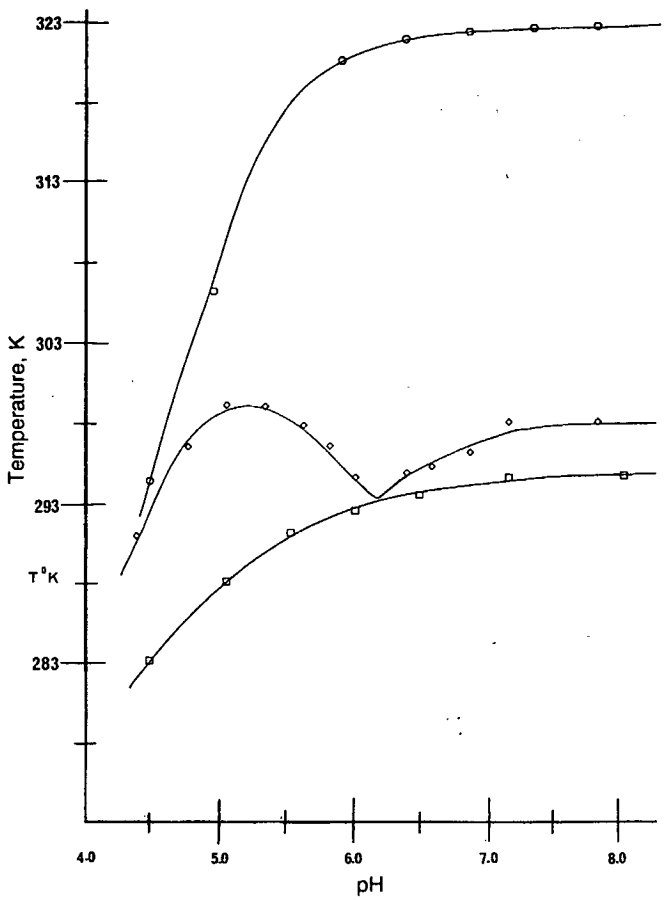


FIG. 3. pH-dependence of ultraviolet melting for d(CGCm^{6}GAATTTGCG) (squares), d(CGCm^{6}GAATTCGCG) (diamonds), and d(CGCGAATTCGCG) control (circles). Ultraviolet melting curves were measured over a wide pH range at 264 nm in aqueous 0.1 M sodium phosphate/1 mM EDTA. Each point on the curve was measured in triplicate.

mispair, which would lead to the formation of a Watson–Crick-like base pair, a resonance form of which is shown in Fig. 4*b*. Thus, the m^{6}G·C base pair displays conformational flexibility. The protonated base pair 4b, which is probably present to some extent at neutral pH, would be expected on structural grounds to be incorporated during replication. Such a base pair has been identified in an NMR study of a mixture of the nucleosides in nonaqueous solvents (21) and in a preliminary report of an x-ray structure of a Z-DNA duplex (33). The occasional occurrence of 4b might explain why the presence of m^{6}G in genomic DNA does not always lead to a mutation.

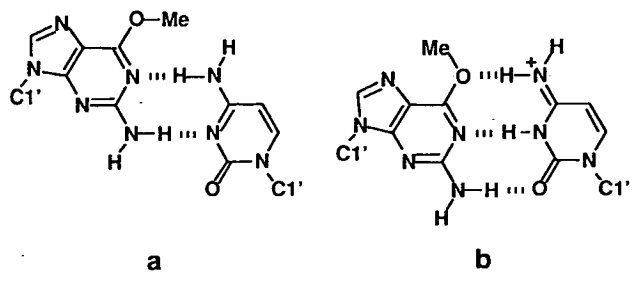


FIG. 4. (*a*) m^{6}G·C wobble base pair. (*b*) Protonated m^{6}G·C$^+$ base pair.

We thank Dr. S. V. Phillips and Dr. W. Summers for valuable guidance in the use of the Rigaku RU200 HB diffractometer, rotating anode, and Nicolet–Siemens X100 area detector system in the Department of Biophysics at Leeds University. This work was funded by a Science and Engineering Research Council Molecular Recognition Initiative grant and a Science and Engineering Research Council CASE Studentship with Shell Research Ltd, U.K.

1. Miller, E. C. & Miller, J. A. (1981) *Cancer* **47**, 2327–2345.
2. Searle, C. E. (1986) *Chem. Br.* **29**, 211–220.
3. Ames, B. N. (1979) *Science* **204**, 587–593.
4. Karran, P. & Marinus, M. G. (1982) *Nature (London)* **296**, 868–869.
5. Saffhill, R., Margison, G. P. & O'Connor, P. (1985) *Biochim. Biophys. Acta* **823**, 111–145.
6. Loechler, E. L., Green, C. L. & Essigman, J. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6271–6275.
7. Mitra, G., Pauly, G. T., Kumar, R., Pei, G. K., Hughes, S. H., Moschel, R. C. & Barbacid, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8650–8654.
8. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. (1982) *Nature (London)* **300**, 149–152.
9. Zarbl, H., Sukumar, S., Arthur, A. V., Martin-Zanka, D. & Barbacid, M. (1985) *Nature (London)* **315**, 382–385.
10. Kuzmich, S., Marky, L. A. & Jones, R. A. (1983) *Nucleic Acids Res.* **11**, 3393–3404.
11. Borowy-Borowski, H. & Chambers, R. W. (1987) *Biochemistry* **26**, 2465–2471.
12. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **287**, 755–758.
13. Kabsch, W. (1988) *J. Applied Cryst.* **21**, 916–924.
14. Sheldrick, G. M. (1976) *Shelx76 System of Computing Programs* (Univ. Cambridge, U.K.).
15. Hendrickson, W. A. & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation, Function and Evolution*, ed. Srinivasan, R. (Pergamon, Oxford), Vol. 1, 43–57.
16. Westhof, E., Dumas, P. & Moras, D. (1985) *J. Mol. Biol.* **184**, 119–145.
17. Jones, T. A. (1978) *J. Applied Cryst.* **11**, 268–272.
18. Gaffney, B. L. & Jones, R. A. (1989) *Biochemistry* **28**, 5881–5889.
19. Gaffney, B. L., Marky, L. A. & Jones, R. A. (1984) *Biochemistry* **23**, 5686–5691.
20. Nagata, C. & Aida, M. (1988) *J. Mol. Struct. (THEOCHEM)* **179**, 451–466.
21. Williams, L. D. & Shaw, B. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1779–1783.
22. Patel, D. J., Shapiro, L., Kozlowski, S., Gaffney, B. L. & Jones, R. A. (1988) *Biochemistry* **25**, 1036–1042.
23. Yamagata, Y., Kohda, K. & Tomita, K. (1988) *Nucleic Acids Res.* **16**, 9307–9321.
24. Parthasarathy, R. & Fridey, S. M. (1986) *Carcinogenesis* **7**, 221–227.
25. Pedersen, L. G., Darden, T. A., Deerfield, D. W., II, Anderson, M. W. & Hoel, D. G. (1988) *Carcinogenesis* **9**, 1553–1562.
26. Snow, E. T., Foote, R. S. & Mitra, S. (1984) *J. Biol. Chem.* **259**, 8095–8100.
27. Chambers, R. W., Sledziewska-Gojska, E., Hirani-Hojatti, S. & Borowy-Borowski, H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7173–7177.
28. Bhanot, O. S. & Ray, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7348–7352.
29. Toorchen, D. & Topal, M. D. (1983) *Carcinogenesis* **4**, 1591–1597.
30. Olsson, M. & Lindahl, T. (1980) *J. Biol. Chem.* **255**, 10569–10571.
31. Demple, B., Jacobsson, A., Olsson, M., Robins, P. & Lindahl, T. (1982) *J. Biol. Chem.* **257**, 13776–13780.
32. Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L. & Jones, R. A. (1986) *Biochemistry* **25**, 1027–1036.
33. Ginell, L., Narendra, N., Jones, R. A. & Berman, H. M. (1990) *Biophys. J.* **57**, 452 (abstr.).

Courses attended

Organic Chemistry Research Seminars, various speakers, Department of Chemistry, Edinburgh University, 1989 - 1992.

Medicinal Chemistry Seminars, Prof. R. Baker and colleagues, Merck, Sharp and Dohme Ltd., 1990 and 1991.

Advances in Organic Chemistry, various speakers. Edinburgh University, 1989 - 1992.

Nucleic Acids and Therapeutics, January 1991, Clearwater Beach, Florida.

Aspects and Applications of NMR Spectroscopy, (various speakers), Edinburgh University.

Royal Society of Chemistry Bio-organic group one day postgraduate symposium, (Warwick University), 1991.

Recognition Studies in Nucleic Acids II, April 1992, Sheffield University.

Departmental German Test, Edinburgh University, 1992.